

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 03:49:50 ; Search time 505.437 Seconds  
(without alignments)  
8094.223 Million cell updates/sec

Title: US-09-966-264D-1

Perfect score: 137

Sequence: 1 attataaaggaaagaaaa.....gtgttgatgttaatt 137

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*

1: em\_estba.\*

2: em\_esthum.\*

3: em\_estin.\*

4: em\_estmu.\*

5: em\_estov.\*

6: em\_estpl.\*

7: em\_estro.\*

8: em\_htc.\*

9: gb\_est1.\*

10: gb\_est2.\*

11: gb\_htc.\*

12: gb\_est3.\*

13: gb\_est4.\*

14: gb\_est5.\*

15: em\_estfun.\*

16: em\_estom.\*

17: em\_gss\_hum.\*

18: em\_gss\_inv.\*

19: em\_gss\_pin.\*

20: em\_gss\_vrt.\*

21: em\_gss\_fun.\*

22: em\_gss\_mam.\*

23: em\_gss\_pro.\*

24: em\_gss\_rod.\*

25: em\_gss\_dhg.\*

26: em\_gss\_vrl.\*

27: em\_gss\_vrl.\*

28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	137	100.0	373	10 AW580404	AW580404 PM2-HT045
C 2	137	100.0	366	14 H89576	H89576 yw28c12.r1
C 3	137	100.0	455	9 AA427831	AA427831 zw49d12.r
C 4	137	100.0	462	14 CB045405	CB045405 NISC_gc10

C 5	137	100.0	473	14 CA389721	CA389721 cs101f03.
C 6	137	100.0	489	28 AQ015185	AQ015185 CIT_HSP-2
C 7	137	100.0	647	10 AW664684	AW664684 hi84e10.x
C 8	137	100.0	696	10 AW950480	AW950480 EST362550
C 9	137	100.0	710	9 AV725574	AV725574 AV725574
C 10	137	100.0	727	12 BG567176	BG567176 602589640
C 11	137	100.0	740	14 CD356811	CD356811 AGENCOURT
C 12	137	100.0	745	14 CD357556	CD357556 AGENCOURT
C 13	137	100.0	745	14 CD357556	CD357556 AGENCOURT
C 14	137	100.0	1121	12 BM546012	BM546012 AGENCOURT
C 15	136	99.3	743	14 CD110842	CD110842 AGENCOURT
C 16	135.4	98.8	375	10 AW607064	AW607064 PM2-HT045
C 17	135.4	98.8	862	14 CB962272	CB962272 AGENCOURT
C 18	135.4	98.8	911	12 BI752714	BI752714 603028327
C 19	120	87.6	352	10 AW385154	AW385154 PM2-HT045
C 20	111.4	81.3	353	10 AW580423	AW580423 PM2-HT045
C 21	107	78.1	631	14 CB465228	CB465228 726457 MA
C 22	100.6	73.4	231	10 BB339107	BB339107 BB339107
C 23	100.6	73.4	390	9 AI324317	AI324317 mq87C07.y
C 24	100.6	73.4	607	10 BE370292	BE370292 601222549
C 25	100.6	73.4	610	9 AA146038	AA146038 mq87C07.f
C 26	100.6	73.4	735	14 CB570707	CB570707 AGENCOURT
C 27	99	72.3	437	11 AK036936	AK036936 Mus muscu
C 28	98.6	72.0	296	10 BB041920	BB041920 BB041920
C 29	97.4	71.1	313	10 BB332959	BB332959 BB332959
C 30	95.8	69.9	210	10 BB171938	BB171938 BB171938
C 31	95.8	69.9	268	10 BB178429	BB178429 BB178429
C 32	95.8	69.9	291	10 BB174122	BB174122 BB174122
C 33	95.8	69.9	329	10 BB085328	BB085328 BB085328
C 34	94.6	69.1	841	10 BF791019	BF791019 602251072
C 35	94.2	68.8	293	9 AV307178	AV307178 AV307178
C 36	92.6	67.6	306	10 BB086855	BB086855 BB086855
C 37	92.6	67.6	332	10 BB235445	BB235445 BB235445
C 38	89.6	65.4	418	13 BY377466	BY377466 BY377466
C 39	86.2	62.9	239	9 AV232779	AV232779 AV232779
C 40	86.2	62.9	293	10 BB307434	BB307434 BB307434
C 41	80.8	59.0	641	9 AI528613	AI528613 mq87C07.x
C 42	78.4	57.2	945	10 BF180441	BF180441 601805231
C 43	72	52.6	955	13 BX419179	BX419179 BX419179
C 44	68.4	49.9	651	13 BU338697	BU338697 603515030
C 45	68.4	49.9	733	13 BU257193	BU257193 603744569

# ALIGNMENTS

RESULT 1	AW580404	373 bp	mRNA	linear	EST 16-MAR-2000
LOCUS	PM2-HT0451-080100-003-h09	HT0451	Homo sapiens	cdna	mRNA sequence.
DEFINITION	AW580404				
ACCESSION	AW580404.1	GI:7255453			
VERSION	AW580404.1	GI:7255453			
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
TITLE	1 (bases 1 to 373)				
JOURNAL	HCGP <a href="http://www.ludwig.org.br/ORESTES">http://www.ludwig.org.br/ORESTES</a> .				
COMMENT	The FAPESP/LICR Human Cancer Genome Project Unpublished (1999) Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: <a href="mailto:asimponeludwig.org.br">asimponeludwig.org.br</a> This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL ( <a href="http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM2&amp;t2=PM2-HT0451-080100-003-h09&amp;l3=2000-01-08&amp;t4=1">http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM2&amp;t2=PM2-HT0451-080100-003-h09&amp;l3=2000-01-08&amp;t4=1</a> )				

Seq primer: puc 18 forward  
High quality sequence stop: 373.  
Location/Qualifiers  
1. .373

# FEATURES

source

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="HT0451"  
/note="Organ: head neck; Vector: puc18; Site 1: SmaI;  
Site 2: SmaI; A mini-library was made by cloning products  
derived from OREGENES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

## ORIGIN

Query Match 100.0%; Score 137; DB 10; Length 373;  
Best Local Similarity 100.0%; Pred. No. 4.1e-25;  
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATTATAAGGAAAAAATAACGCAATGCGACAACTGCTGAAGCTGTGAAGCTGTGAAGCTGTGT 60  
DB 369 ATTATAAGGAAAAAATAACGCAATGCGACAACTGCTGAAGCTGTGAAGCTGTGTGAAGCTGTGT 310  
QY 61 GCACAAATTATCAGGACACCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCCCTG 120  
DB 309 GCACAAATTATCAGGACACCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCCCTG 250  
QY 121 TTGTGTTTAATTAATT 137  
DB 249 TTGTGTTTAATTAATT 233

## RESULT 2

H89576 396 bp mRNA linear EST 28-NOV-1995  
LOCUS Yw28c12.r1 Morton Fetal Cochlea Homo sapiens cDNA clone  
DEFINITION IMAGE:253558 5' similar to gb:M18533 DYSTROPHIN (HUMAN); mRNA  
sequence.

ACCESSION H89576.1 GI:1079922

VERSION H89576

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 396)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## AUTHORS

Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,  
Hawkins, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,  
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
Trevisan, K., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.,  
and Marra, M.

Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)

## JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: Wilton RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

High quality sequence stops: 300

Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1088 Std Error: 0.00

Seq primer: M13RP1

High quality sequence stop: 300.

# FEATURES

source

Location/Qualifiers  
1. .396  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="GDB:3891011"  
/db\_xref="taxon:9606"  
/clone="IMAGE:253558"  
/tissue\_type="cochlea"  
/dev\_stage="16-22 week fetus"  
/lab\_host="SOLR cells (kanamycin resistant)"  
/clone\_lib="Morton Fetal Cochlea"  
/note="Organ: ear; Vector: pBluescript SK-; Site 1: EcoRI;  
Site 2: XhoI; Reference: Genomics 23, 42-50 (1994) Cloned  
unidirectionally. Primer: Oligo dt. Fetal cochlea, normal.  
37% of inserts <0.5 kb, 56% 0.5-1.0 kb, 7% >1 kb. Uni-ZAP  
XR Vector. Library constructed by N. Robertson, C. Morton.  
-5' adaptor sequence: 5' GAATTCGACGAG 3' ~3' adaptor  
sequence: 5' CTCGAGTTTTTTTTTTTTTT 3'

## ORIGIN

Query Match 100.0%; Score 137; DB 14; Length 396;  
Best Local Similarity 100.0%; Pred. No. 4.1e-25;  
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATTATAAGGAAAAAATAACGCAATGCGACAACTGCTGAAGCTGTGAAGCTGTGTGAAGCTGTGT 60  
DB 264 ATTATAAGGAAAAAATAACGCAATGCGACAACTGCTGAAGCTGTGTGAAGCTGTGTGAAGCTGTGT 205  
QY 61 GCACAAATTATCAGGACACCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCCCTG 120  
DB 204 GCACAAATTATCAGGACACCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCCCTG 145  
QY 121 TTGTGTTTAATTAATT 137  
DB 144 TTGTGTTTAATTAATT 128

## RESULT 3

AA427831/c

LOCUS

DEFINITION

AA427831 455 bp mRNA linear EST 16-OCT-1997

IMAGE:773399 5' similar to gb:M18533 DYSTROPHIN (HUMAN); mRNA

sequence.

ACCESSION AA427831

VERSION AA427831.1 GI:2111628

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 455)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Hillier, L., Allen, M., Bowles, L., Dubuc, T., Geisel, G., Jost, S.,

Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J.,

Moore, B., Schellenberg, K., Stepien, M., Tan, F., Theising, B.,

White, Y., Wylie, T., Waterston, R. and Wilson, R.

WashU-Merck EST Project 1997

Unpublished (1997)

Contact: Wilton RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 391.

Location/Qualifiers

1. .455

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:773399"

# FEATURES

source

/dev\_stage="8-9 weeks"  
 /lab\_host="DH108"  
 /clone\_lib="Soares total\_fetus Nb2HF8\_9w"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
 was prepared from mRNA obtained from pooled 8-9 week  
 (total) fetus material with a Not I - oligo(dT) primer [5'  
 TGTTACCATCTGAGTGGAGCGCGCTTAATTTTCTTTTCTTTT 3'].  
 Double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT73 vector. Library  
 went through one round of normalization, and was  
 constructed by Bento Soares and M. Fatima Bonaldo. "

## ORIGIN

Query Match 100.0%; Score 137; DB 9; Length 455;  
 Best Local Similarity 100.0%; Pred. No. 4e-25;  
 Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATAAGGAAAAAAGAAATAACGCAATGGCAAGTGGTGAAGTGTGAACCTCAGGTGT 60  
 Db 155 ATTATAAGGAAAAAAGAAATAACGCAATGGCAAGTGGTGAAGTGTGAACCTCAGGTGT 96

QY 61 GCACAAATTCAGGAACACCCCAACCAACCAAGTGGTAGAAATAGCATGAGAACCCGTG 120  
 Db 95 GCACAAATTCAGGAACACCCCAACCAACCAAGTGGTAGAAATAGCATGAGAACCCGTG 36

QY 121 TTGTGATGTTAATTAATT 137  
 Db 35 TTGTGATGTTAATTAATT 19

## RESULT 4

CB045405/c  
 LOCUS  
 DEFINITION NISC gc10c01.v1 NCI\_CGAP\_Col17 Homo sapiens cDNA clone IMAGE:3218281  
 5', mRNA sequence.

ACCESSION CB045405  
 VERSION CB045405.1 GI:27783692  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 462)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs@mail.nih.gov](mailto:cgapbs@mail.nih.gov)

CDNA Library Preparation:  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL  
 DNA Sequencing by: National Institutes of Health Intramural  
 Sequencing Center (NISC)  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[infoimage.llnl.gov](http://infoimage.llnl.gov)  
 Plate: LLAW7870 row: E column: 2  
 Seq primer: M13RPI reverse primer (ABI).

## FEATURES

source  
 1..462  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3218281"  
 /tissue\_type="juvenile granulosa tumor"  
 /lab\_host="DH108"  
 /clone\_lib="NCI CGAP Col17"  
 /note="Organ: colon; Vector: pCMV-SPORT6; Site 1: SalI;  
 Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
 Library constructed by Life Technologies."

## ORIGIN

Query Match 100.0%; Score 137; DB 14; Length 462;  
 Best Local Similarity 100.0%; Pred. No. 4e-25;  
 Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATAAGGAAAAAAGAAATAACGCAATGGCAAGTGGTGAAGTGTGAACCTCAGGTGT 60  
 Db 432 ATTATAAGGAAAAAAGAAATAACGCAATGGCAAGTGGTGAAGTGTGAACCTCAGGTGT 373

QY 61 GCACAAATTCAGGAACACCCCAACCAACCAAGTGGTAGAAATAGCATGAGAACCCGTG 120  
 Db 372 GCACAAATTCAGGAACACCCCAACCAACCAAGTGGTAGAAATAGCATGAGAACCCGTG 313

QY 121 TTGTGATGTTAATTAATT 137  
 Db 312 TTGTGATGTTAATTAATT 296

## RESULT 5

CA389721/c  
 LOCUS  
 DEFINITION cs101f03.y3 Human Retinal pigment epithelium/choroid cDNA  
 (Un-normalized, unamplified): cs Homo sapiens choroid clone cs101f03  
 5', mRNA sequence.

ACCESSION CA389721  
 VERSION CA389721.1 GI:24720152  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 473)  
 Touchman,J.W., Bouffard,G., Smith,D. and Peterson,K.  
 Wistow,G., Bernstein,S.L., Wyatt,M.K., Farris,R.N., Behal,A.,  
 Expressed sequence tag analysis of human RPE/choroid for the  
 NIEBANK Project: Over 6000 non-redundant transcripts, novel genes  
 and splice variants

JOURNAL Mol. Vis. 8 (4), 205-220 (2002)  
 MEDLINE 22103460  
 PUBMED 12107410  
 COMMENT Contact: Wistow G  
 Section on Molecular Structure and Function  
 National Eye Institute  
 6/331, NIH, Bethesda, MD 20892-2740, USA  
 Tel: 301 402 3452  
 Fax: 301 496 0078  
 Email: [graeme@helix.nih.gov](mailto:graeme@helix.nih.gov)  
 Plate: 101 row: f column: 03  
 Seq primer: M13RPI reverse primer (ABI).

## FEATURES

source  
 1..473  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="cs101f03"  
 /tissue\_type="RPE/choroid"  
 /dev\_stage="Adult"  
 /lab\_host="EMDH108"  
 /clone\_lib="Human Retinal pigment epithelium/choroid cDNA  
 (Un-normalized, unamplified): cs"  
 /notes="Organ: Eye; Vector: pCMVSPORT6; Two different donor  
 eyes (75-80 years old) yielded approximately 600 mg of  
 dissected RPE/choroid tissue. This in turn yielded 340 ug  
 of total RNA and 7 ug of mRNA. A directionally cloned cDNA  
 library in the pCMVSPORT6 vector was constructed at Life  
 Technologies (Rockville, MD; now part of Invitrogen Corp),  
 essentially following the protocols of the SuperScript  
 Plasmid System (Invitrogen Corp).  
 <<http://www.invitrogen.com/>>.  
 The library code  
 designation was cs. For this library, cDNA inserts were  
 cloned into the NotI/MluI sites of the vector. EST  
 analysis was performed on the unamplified library at the  
 NIH Intramural Sequencing Center (NISC)."

```

ORIGIN
Query Match      100.0%; Score 137; DB 14; Length 473;
Best Local Similarity 100.0%; Pred. No. 3.9e-25;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  ATTATAAGGAAAAAGAAATAACGCCAATGTGACAAAGTGGTGAAGCTGTGAAGCTCAGGTGT 60
Db      231  ATTATAAGGAAAAAGAAATAACGCCAATGTGACAAAGTGGTGAAGCTGTGAAGCTCAGGTGT 172

QY      61  GCACAAATTATCAGGAACACCCCAAAACCAAAAGTGGGTAGAGTAAGTATGAGTAAGAGCCGGT 120
Db      171  GCACAAATTATCAGGAACACCCCAAAACCAAAAGTGGGTAGAGTAAGTATGAGTAAGAGCCGGT 112

QY      121  TTTGATGTTAAATTAATT 137
Db      111  TTTGATGTTAAATTAATT 95

RESULT 6
AQ015185/c
LOCUS      AQ015185
DEFINITION  CIT-HSP-2310F4.TR CIT-HSP Homo sapiens genomic clone 2310F4,
            genomic survey sequence.
ACCESSION  AQ015185
VERSION    AQ015185.1 GI:3193921
KEYWORDS   GSS.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 489)
AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
            Simon,M. and Venter,J.C.
TITLE      Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building (1998)
JOURNAL    Unpublished (1998)
COMMENT    Other GSSs: CIT-HSP-2310F4.TF
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mcadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends
FEATURES   Location/Qualifiers
            1..489
             /organism="Homo sapiens"
             /mol_type="genomic DNA"
             /db_xref="taxon:9606"
             /clone="2310F4"
             /sex="Male"
             /cell_type="Sperm"
             /clone_lib="CIT-HSP"
             /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
             HindIII"

ORIGIN
Query Match      100.0%; Score 137; DB 28; Length 489;
Best Local Similarity 100.0%; Pred. No. 3.9e-25;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  ATTATAAGGAAAAAGAAATAACGCCAATGTGACAAAGTGGTGAAGCTGTGAAGCTCAGGTGT 60
Db      358  ATTATAAGGAAAAAGAAATAACGCCAATGTGACAAAGTGGTGAAGCTGTGAAGCTCAGGTGT 299

QY      61  GCACAAATTATCAGGAACACCCCAAAACCAAAAGTGGGTAGAGTAAGTATGAGTAAGAGCCGGT 120

```



Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
Source
Location/Qualifiers
1..710
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HTCETD10"
/tissue_type="Hypothalamus"
/dev_stage="Adult"
/lab_host="SOL3"
/clone_lib="HTC"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
ORIGIN
Query Match 100.0%; Score 137; DB 9; Length 710;
Best Local Similarity 100.0%; Pred.No.3.7e-25;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATTATAAGGAAAAGAAATTAACCAATGCGAAGTGTGAAGCTGTCAACTCAGGTGT 60
DB 396 ATTATAAGGAAAAGAAATTAACGCAATGGCAAGTGTGAAGCTGTCAACTCAGGTGT 337
QY 61 GCACATTATCATGGACACCCCAAAACCACCAAGTCAGGTAGTAATACATGAGNAGCCGTC 120
DB 336 GCACAATTATCATGGACACCCCAAAACCACCAAGTCAGGTAGTAATACATGAGNAGCCGTC 120
QY 121 TTTTGATGTTAATTAATT 137
DB 276 TTTTGATGTTAATTAATT 260
RESULT 10
BGS67176/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
1 (bases 1 to 727)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
cDNA library Preparation: CLONETECH Laboratories, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
Plate: LLCML583 row: P column: 10
High quality sequence stop: 694.
Location/Qualifiers
1..727
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4723833"
/lab_host="DH10B (TI phase-resistant)"
/clone_lib="NH_MGC_76"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
sfiI (ggcgctcgcc); Site_2: SfiI (ggccattatgcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCCATTATGCCC-3' and 3' adaptor sequence:

Note: this is a NIH MGC Library."

550	ATTATTAAGGAAAAAGAAAAATTAACGCAATGGACAAGTCGTTGAAGCTGTGAACTCAGGTGT	491
Db		
61	GCACAAATTATCAGGAACACCCCAAAACCACAAAGTCAGGTAGAGAAATAGCATGAGAAACCGGTG	120
Qy		
490	GCACAAATTATCAGGAACACCCCAAAACCACAAAGTCAGGTAGAGAAATAGCATGAGAAACCGGTG	431
Db		
121	TTTGCATGTTAATTAAAT	137
Qy		
430	TTTGCATGTTAATTAAAT	414
Db		

LOCUS	CD357556	745 bp	mRNA	linear	EST 29-MAY-2003
DEFINITION	AGENCOURT 14253482	NIH_MGC_187	Homo sapiens	CDNA clone	
IMAGE	30402369	5'	mrna	sequence.	

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Mammalia; Eutheria; Primates; Catarrhini, Hominoidea; Homo.  
1 (bases 1 to 745)  
NIH-MGC <http://mgc.nhl.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Daniela S. Gerhard, Ph.D.

CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
CDNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: NDCMi84 row: j column: 10  
High quality sequence stop: 460.

```

/clones="IMAGE:30402369"
/lab host="DH10B (T1 phage-resistant)"
/clone lib="NIH_MGC 187"
/note="Organ: Blood vessels - aorta, basilar and artery;
Vector: pDNR-LIB; Site_1: SfiI (ggccattatggcc); Site_2:
SfiI (ggccgctctggcc); 5' and 3' adaptors were used in
cloning as follows: 5' adaptor sequence:
5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTATAGACCCGCGCCGACATG-DT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.4 kb
(range 0.5-4.0 kb). 14/15 colonies contained inserts by
PCR. This library was enriched for full-length clones and
was constructed by Clontech Laboratories (Palo Alto, CA).
Note: this is a NIH MGC library."

```

Query Match 100.0%; Score 137; DB 14; Length 745;  
Best Local Similarity 100.0%; Pred. No. 3.6e-25;

**PCR.** This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA) (range 0.5–4.0 kb). 14/15 colonies contained inserts by PCR. This library was enriched for full-length clones and

Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATAAGGAAAAAAGAAATAACCAATGCAAGTGGTGAAGCTGTGAACCTCAGGTGT 60  
 DB 260 ATTATAAGGAAAAAAGAAATAACCAATGCAAGTGGTGAAGCTGTGAACCTCAGGTGT 201

QY 61 GCACAATTATCAGGAACACCCCAACCAACCAAGTGGTAGAATAAGCATGAGAGCCGCTG 120  
 DB 200 GCACAATTATCAGGAACACCCCAACCAACCAAGTGGTAGAATAAGCATGAGAGCCGCTG 141

QY 121 TTGTGATGTTAATTAATT 137  
 DB 140 TTGTGATGTTAATTAATT 124

RESULT 13  
 BM546012/c  
 LOCUS  
 DEFINITION BM546012 1121 bp mRNA linear EST 20-FEB-2002  
 AGENCOURT 6505286 NIH\_MGC\_125 Homo sapiens cDNA clone IMAGE:5588419  
 5', mRNA sequence.

ACCESSION BM546012  
 VERSION BM546012.1 GI:18778623  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 1121)  
 NIH-MGC http://mgc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 JOURNAL  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Invitrogen  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLN12359 row: h column: 20  
 High quality sequence start: 50  
 High quality sequence stop: 764.

FEATURES  
 Location/Qualifiers  
 1..1121  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5588419"  
 /lab\_host="DH10B"  
 /clone\_lib="NIH MGC 125"  
 /note="Organ: ovary (pool of 3); Vector: pCMV-SPORT6;  
 Site 1: EcoRV (destroyed); Site 2: NotI; RNA source pool  
 of three ovaries, from females ranging in age from 38 to  
 49 yo. Library is oligo-dT primed and directionally cloned  
 (EcoRV site is destroyed upon cloning). Average insert  
 size 2.1 kb, insert size range 1-3.5 kb. Library is  
 normalized and enriched for full-length clones and was  
 constructed by C. Gruber (Invitrogen). Research Genetics  
 tracking code 036."

ORIGIN  
 Query Match 100.0%; Score 137; DB 12; Length 1121;  
 Best Local Similarity 100.0%; Pred. No. 3.4e-25;  
 Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATAAGGAAAAAAGAAATAACCAATGCAAGTGGTGAAGCTGTGAACCTCAGGTGT 60  
 DB 514 ATTATAAGGAAAAAAGAAATAACCAATGCAAGTGGTGAAGCTGTGAACCTCAGGTGT 455

QY 61 GCACAATTATCAGGAACACCCCAACCAACCAAGTGGTAGAATAAGCATGAGAGCCGCTG 120  
 DB 454 GCACAATTATCAGGAACACCCCAACCAACCAAGTGGTAGAATAAGCATGAGAGCCGCTG 395

QY 121 TTGTGATGTTAATTAATT 137  
 DB 394 TTGTGATGTTAATTAATT 378

RESULT 14  
 CD110642/c  
 LOCUS  
 DEFINITION CD110642 743 bp mRNA linear EST 15-MAY-2003  
 AGENCOURT 13995397 NIH\_MGC\_187 Homo sapiens cDNA clone  
 IMAGE:30373580 5', mRNA sequence.

ACCESSION CD110642  
 VERSION CD110642.1 GI:30754851  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 743)  
 NIH-MGC http://mgc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 JOURNAL  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits  
 cDNA Library Preparation: CLONTECH Laboratories, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: NDCM167 row: j column: 21  
 High quality sequence stop: 578.

FEATURES  
 Location/Qualifiers  
 1..743  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:30373580"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /clone\_lib="NIH MGC 187"  
 /note="Organ: Blood vessels - aorta, basilar and artery;  
 Vector: pMNR-Lib; Site 1: SfiI (ggccattatggcc); Site 2:  
 SfiI (ggccgctcgcc); 5' and 3' adaptors  
 cloning as follows: 5' adaptor sequence:  
 5'-CAGCGCCATTATGCCC-3' and 3' adaptor sequence:  
 5'-ATTCTAGAGCCGCGCGCCGACATG-dt(30)BN-3' (where B = A,  
 C, or G and N = A, C, G, or T). Average insert size 1.4 kb  
 (range 0.5-4.0 kb). 14/15 colonies contained inserts by  
 PCR. This library was enriched for full-length clones and  
 was constructed by Clontech Laboratories (Palo Alto, CA).  
 Note: this is a NIH\_MGC Library."

ORIGIN  
 Query Match 99.3%; Score 136; DB 14; Length 743;  
 Best Local Similarity 100.0%; Pred. No. 6.6e-25;  
 Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATAAGGAAAAAAGAAATAACCAATGCAAGTGGTGAAGCTGTGAACCTCAGGTGT 60  
 DB 138 ATTATAAGGAAAAAAGAAATAACCAATGCAAGTGGTGAAGCTGTGAACCTCAGGTGT 79

QY 61 GCACAATTATCAGGAACACCCCAACCAACCAAGTGGTAGAATAAGCATGAGAGCCGCTG 120  
 DB 78 GCACAATTATCAGGAACACCCCAACCAACCAAGTGGTAGAATAAGCATGAGAGCCGCTG 19

QY 121 TTGTGATGTTAATTAATT 136  
 DB 18 TTGTGATGTTAATTAATT 3

RESULT 15  
 AW607064/c

LOCUS AW607064 375 bp mRNA linear EST 23-MAR-2000  
 DEFINITION PM2-HT0451-170100-004-h12 HT0451 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION AW607064  
 VERSION AW607064.1 GI:7311805  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 375)  
 HCGF <http://www.ludwig.org.br/ORESTES>.  
 The FAPESP/LICR Human Cancer Genome Project  
 Unpublished (1999)  
 CONTACT: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM2&t2=PM2-HT0451-170100-004-h12&t3=2000-01-17&t4=1>)  
 Seq primer: puc 18 forward  
 High quality sequence stop: 375.

## FEATURES

Location/Qualifiers  
 1..375  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /dev\_stage="Adult"  
 /clone\_lib="HT0451"  
 /note="Organ: head neck; Vector: puc18; Site 1: Smal;  
 Site 2: Smal; A mini-library was made by cloning products  
 derived from ORESTES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the pUC 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."

## ORIGIN

Query Match 98.8%; Score 135.4; DB 10; Length 375;  
 Best Local Similarity 99.3%; Pred. No. 1.1e-24;  
 Matches 136; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 ATTATAAGGAAAAAGAAATACGCAATGGACAGCTGTGAAGCTGTGAAGCTCAGGTGT 60  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 371 ATTATAAGGAAAAAGAAATACGCAATGGACAGCTGTGAAGCTGTGAAGCTCAGGTGT 312  
 QY 61 GCACATTATCAGGAACACCCCAAAACCAAGTAGGTAAGTAATAGCATGAGAGCCGTG 120  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 311 GCACATTATCTGGAACACCCCAAAACCAAGTAGGTAAGTAATAGCATGAGAGCCGTG 252  
 QY 121 TTGATGTTTAATTAAT 137  
 Db ||||||||||||||||  
 251 TTGATGTTTAATTAAT 235

Search completed: April 6, 2004, 14:12:02  
 Job time : 511.437 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 02:56:20 ; Search time 64.812 Seconds  
(without alignments)

8979.866 Million cell updates/sec

Title: US-09-966-264D-1

Perfect score: 137

Sequence: 1 attataaaggaaaagaaaa.....gtgtttgatgtaattaatt 137

Scoring table: IDENTITY NUC

Gapop 10\_0 , Gapext 1.0

Searched: 337863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747736

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N Geneseq 29Jan04:\*

1: Geneseqn1980a:\*

2: Geneseqn1990a:\*

3: Geneseqn2000a:\*

4: Geneseqn2001a:\*

5: Geneseqn2002a:\*

6: Geneseqn2003a:\*

7: Geneseqn2004a:\*

8: Geneseqn2005a:\*

9: Geneseqn2006a:\*

10: Geneseqn2007a:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	137	100.0	137	6	ABK86496 Human Apo
C 2	137	100.0	158	6	ABK86469 Human Apo
C 3	137	100.0	200	6	ABK86468 Human Apo
C 4	137	100.0	455	9	ADD32504 Human mit
5	137	100.0	996	6	ABK86497 Human Apo
6	137	100.0	1230	6	ABK86462 Human Apo
7	137	100.0	1234	6	ABK86463 Human Apo
C 8	137	100.0	2691	6	ABK81996 Human Apo
C 9	137	100.0	8689	6	ABK82000 DNA encod
C 10	137	100.0	11443	6	ABK82002 DNA encod
C 11	137	100.0	13957	6	ABK81959 CDNA enco
C 12	137	100.0	13957	6	ABK10904 Human bre
C 13	137	100.0	13957	6	ABK95786 Gene #228
C 14	137	100.0	13957	6	ABK69900 Human dys
C 15	126	92.0	13977	6	ABK70403 Human bon
C 16	107	78.1	959	6	ABK74601 Bovine em
C 17	100.6	73.4	13815	2	ABK18885 Mus muscu
C 18	100.6	73.4	13815	6	ABK81960 CDNA enco
C 19	100.6	73.4	13815	6	ABK199799 Mouse isc
C 20	100.6	73.4	13815	2	ABK27558 Shuttle v
C 21	57.2	41.8	108	6	ABK86467 Human apo
C 22	41	29.9	60	6	ABK86471 Human apo
C 23	33.6	24.5	1863	9	ADK60952 Human gen

C 24	31.8	23.2	6428	4	AAK78562	Human imm
C 25	31.8	23.2	96591	9	ADC85301	Mouse Sos
C 26	31.8	23.2	96592	8	ADA02822	Human SOS
C 27	31.8	23.2	96592	6	ADB72560	Human SOS
C 28	31.6	23.1	2000	6	ABZ17261	Arabidops
C 29	31.2	22.8	1977	6	ABU62657	Colin ade
C 30	31.2	22.8	1977	6	ABT10871	Human bre
C 31	31.2	22.8	1977	6	ABK83769	Human CDN
C 32	31.2	22.8	1979	2	AAT87015	Human FC-
C 33	31.2	22.8	1990	1	AAN92451	cDNA inse
C 34	31.2	22.8	2461	6	AAS94845	Human DNA
C 35	30.8	22.5	47319	4	AAK72230	Human imm
C 36	30.8	22.5	47319	4	AAK64813	Human imm
C 37	30.8	22.5	236303	4	AAS11614	Human gen
C 38	30.6	22.3	538	6	ABL78454	Human Ova
C 39	30.6	22.3	539	6	ABK62246	Rat sequ
C 40	30.6	22.3	539	7	ABT40428	Toxicity
C 41	30.6	22.3	539	9	ADB55325	Toxicity
C 42	30.6	22.3	539	9	ADB49833	Primary r
C 43	30.6	22.3	6644	2	AAX33181	Base sequ
C 44	30.6	22.3	7372	2	AAX33182	Base sequ
C 45	30.6	22.3	7797	2	AAX33180	Cowpox vi

#### ALIGNMENTS

RESULT 1  
ABK86496  
ID ABK86496 standard; DNA; 137 BP.  
XX AC ABK86496;  
XX DT 27-AUG-2002 (first entry)  
XX DE Human Apo-dystrophin-4 inversion sequence.  
XX KW Human; ds; apo-dystrophin-4; inversion sequence; gene therapy;  
XX KM protein truncation; muscular dystrophy; leukaemia.  
XX OS Homo sapiens.  
XX FH Key  
XX CDS  
XX Location/Qualifiers  
3. .137  
/tag= a  
/product= "apo-dystrophin-4 peptide appearing as  
AAU98738"  
/partial  
/note= "No start or stop codon shown"  
/transl\_except= (pos:21..23,aa:Xaa)  
/transl\_except= (pos:48..50,aa:Xaa)  
/transl\_except= (pos:93..95,aa:Xaa)  
/transl\_except= (pos:123..125,aa:Xaa)  
/transl\_except= (pos:129..131,aa:Xaa)  
/note= "Xaa= unknown, encoded by in frame stop codon"

GB2368064-A.  
24-APR-2002.  
16-JAN-2001; 2001GB-00001124.  
30-SEP-2000; 2000US-0237079P.  
(IMCR ) IMPERIAL CANCER RES TECHNOLOGY LTD.  
(BARB/) BARBER E.  
Barber E;  
WPI; 2002-429042/46.  
P-PSDB; AAU98737.  
New human regulatory polynucleotide, useful for treating disorders

PT associated with protein truncation, particularly muscular dystrophy, and  
 PT related peptides and antibodies.  
 XX Claim 1; Page 169; 222pp; English.

XX The invention relates to a polynucleotide (I) comprising, or consisting  
 CC of, apo-dystrophin-4 inversion sequence appearing as ABK86496, or its  
 CC functional equivalents (e.g. the apo-dystrophin-4 cDNA sequence appearing  
 CC as ABK86497). Also included are polynucleotides that hybridise to either  
 CC strand of (I), a vector containing (I), a cell containing (I) or the  
 CC vector, proteins and peptides encoded by (I), a protein homologous with  
 CC human dystrophin that is expressed on cell surfaces in vivo antibodies  
 CC (Ab) specific for the protein and method of screening for leukemia cells  
 CC The apo-dystrophin-4 inversion sequence is a regulatory element that  
 CC controls expression (transcription and translation) of associated DNA,  
 CC and may allow read-through of stop codons. The apo-dystrophin-4 inversion  
 CC of proteins, particularly muscular dystrophy and also leukaemia, but more  
 CC generally (I) is a regulatory sequence used to control expression of any  
 CC attached gene. Analysis of DNA for (I), or detection of proteins (II)  
 CC encoded by (I), can be used to screen for leukaemic cells and related  
 CC diseases. Antibodies raised against (II) can be used therapeutically, to  
 CC inhibit (II) activity, also to detect (II) in screening assays. The  
 CC present sequence is the apo-dystrophin-4 inversion sequence and upstream  
 CC genomic region

XX Sequence 137 BP; 56 A; 19 C; 32 G; 30 T; 0 U; 0 Other;

Query Match 100.0%; Score 137; DB 6; Length 137;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-33;  
 Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATAAGGAAAAAGAAATAACGCAATGGACAGTGGTGAAGCTGTGAAGTCACTCAGGTGT 60

Db 1 ATTATAAGGAAAAAGAAATAACGCAATGGACAGTGGTGAAGCTGTGAAGTCACTCAGGTGT 60

QY 61 GCACAAATTATCAGGAACACCCCAACCAAGTGAAGTGTAGAAATAGCATGAGAACCCGTG 120

Db 61 GCACAAATTATCAGGAACACCCCAACCAAGTGAAGTGTAGAAATAGCATGAGAACCCGTG 120

QY 121 TTGTGTTTAATTAATT 137

Db 121 TTGTGTTTAATTAATT 137

# RESULT 2

ABK86469/C  
 ID ABK86469 standard; cDNA; 158 BP.

XX ABK86469;

XX 27-AUG-2002 (first entry)

XX Human apo-dystrophin-4 cDNA fragment with inversion breakpoint #2.

XX Human; ss; apo-dystrophin-4; inversion sequence; gene therapy;  
 KW protein truncation; muscular dystrophy; leukaemia; dystrophin.

XX OS Homo sapiens.

XX Key Location/Qualifiers

FT misc\_recomb 137

FT /\*tag= a  
 FT /label= inversion\_breakpoint

XX GB2368064-A.

XX 24-APR-2002.

XX 16-JAN-2001; 2001GB-00001124.

XX 30-SEP-2000; 2000US-0237079P.

(IMCR ) IMPERIAL CANCER RES TECHNOLOGY LTD.

(BARB/) BARBER E.

Barber E;

WPI; 2002-429042/46.

New human regulatory polynucleotide, useful for treating disorders  
 associated with protein truncation, particularly muscular dystrophy, and  
 related peptides and antibodies.

Disclosure; Fig 16B; 222pp; English.

The invention relates to a polynucleotide (I) comprising, or consisting  
 of, apo-dystrophin-4 inversion sequence appearing as ABK86496, or its  
 functional equivalents (e.g. the apo-dystrophin-4 cDNA sequence appearing  
 as ABK86497). Also included are polynucleotides that hybridise to either  
 strand of (I), a vector containing (I), a cell containing (I) or the  
 vector, proteins and peptides encoded by (I), a protein homologous with  
 human dystrophin that is expressed on cell surfaces in vivo antibodies  
 (Ab) specific for the protein and method of screening for leukemia cells  
 by analysing DNA for presence of (I) or by detecting presence of (II).  
 The apo-dystrophin-4 inversion sequence is a regulatory element that  
 controls expression (transcription and translation) of associated DNA,  
 and may allow read-through of stop codons. The apo-dystrophin-4 inversion  
 sequence is used in gene therapy of diseases associated with truncation  
 of proteins, particularly muscular dystrophy and also leukaemia, but more  
 generally (I) is a regulatory sequence used to control expression of any  
 attached gene. Analysis of DNA for (I), or detection of proteins (II)  
 encoded by (I), can be used to screen for leukaemic cells and related  
 diseases. Antibodies raised against (II) can be used therapeutically, to  
 inhibit (II) activity, also to detect (II) in screening assays. The  
 present sequence is a apo-dystrophin-4 cDNA fragment showing an inversion  
 breakpoint (recombination signal sequence) similar to that in with human  
 dystrophin

XX Sequence 158 BP; 35 A; 35 C; 20 G; 68 T; 0 U; 0 Other;

Query Match 100.0%; Score 137; DB 6; Length 158;

Best Local Similarity 100.0%; Pred. No. 1.7e-33;

Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATAAGGAAAAAGAAATAACGCAATGGACAGTGGTGAAGCTGTGAAGTCACTCAGGTGT 60

Db 137 ATTATAAGGAAAAAGAAATAACGCAATGGACAGTGGTGAAGCTGTGAAGTCACTCAGGTGT 78

QY 61 GCACAAATTATCAGGAACACCCCAACCAAGTGAAGTGTAGAAATAGCATGAGAACCCGTG 120

Db 77 GCACAAATTATCAGGAACACCCCAACCAAGTGAAGTGTAGAAATAGCATGAGAACCCGTG 18

QY 121 TTGTGTTTAATTAATT 137

Db 17 TTGTGTTTAATTAATT 1

# RESULT 3

ABK86468/C

ID ABK86468 standard; DNA; 200 BP.

XX ABK86468;

XX 27-AUG-2002 (first entry)

XX Human dystrophin genomic DNA with inversion breakpoint #2.

XX Human; ds; apo-dystrophin-4; inversion sequence; gene therapy;  
 KW protein truncation; muscular dystrophy; leukaemia; dystrophin.

XX OS Homo sapiens.

XX Key Location/Qualifiers

FT misc\_recomb 42

```

FT FT      /*tag= a
FT FT      /label= Inversion_breakpoint
FT FT      179
FT FT      /*tag= b
FT FT      /label= Inversion_breakpoint
XX
XX GB2368064-A.
XX
XX 24-APR-2002.
XX
XX 16-JAN-2001; 2001GB-00001124.
XX
XX 30-SEP-2000; 2000US-0237079P.
XX
XX (IMCR ) IMPERIAL CANCER RES TECHNOLOGY LTD.
XX PA (BARB/) BARBER E.
XX
XX Barber E;
XX
XX WPI; 2002-429042/46.
XX
XX New human regulatory polynucleotide, useful for treating disorders
XX PT associated with protein truncation, particularly muscular dystrophy, and
XX PT related peptides and antibodies.
XX
XX Disclosure; Fig 16B; 222pp; English.
XX
XX The invention relates to a polynucleotide (I) comprising, or consisting
XX CC of, apo-dystrophin-4 inversion sequence appearing as ABK86496, or its
XX CC functional equivalents (e.g. the apo-dystrophin-4 cDNA sequence appearing
XX CC as ABK86497). Also included are polynucleotides that hybridise to either
XX CC strand of (I), a vector containing (I), a cell containing (I) or the
XX CC vector, proteins and peptides encoded by (I), a protein homologous with
XX CC human dystrophin that is expressed on cell surfaces in vivo antibodies
XX CC (Ab) specific for the protein and method of screening for leukemia cells
XX CC by analysing DNA for presence of (I) or by detecting presence of (II).
XX CC The apo-dystrophin-4 inversion sequence is a regulatory element that
XX CC controls expression (transcription and translation) of associated DNA,
XX CC and may allow read-through of stop codons. The apo-dystrophin-4 inversion
XX CC sequence is used in gene therapy of diseases associated with truncation
XX CC of proteins, particularly muscular dystrophy and also leukaemia, but more
XX CC generally (I) is a regulatory sequence used to control expression of any
XX CC attached gene. Analysis of DNA for (I), or detection of proteins (II)
XX CC encoded by (I), can be used to screen for leukaemic cells and related
XX CC diseases. Antibodies raised against (II) can be used therapeutically, to
XX CC inhibit (II) activity, also to detect (II) in screening assays. The
XX CC present sequence is a human dystrophin genomic DNA fragment showing
XX CC inversion breakpoints (recombination signal sequence) similar to that in
XX CC apo-dystrophin-4
XX
XX Sequence 200 BP; 44 A; 40 C; 28 G; 88 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 137; DB 6; Length 200;
XX Best Local Similarity 100.0%; Pred. No. 1.8e-33;
XX Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 ATTATAAGGAAAAAGAAAAATACCAATGGCAAGTGTGTGAAGCTGTGAAGCTCAGGTGT 60
XX DB |||||
XX 179 ATTATAAGGAAAAAGAAAAATACCAATGGCAAGTGTGTGAAGCTGTGAAGCTCAGGTGT 120
XX
XX 61 GCACATTTATCAGGACACCCCAACCAACCAAGTGTAGTAAATAGCATGAGACCCGTG 120
XX QY |||||
XX 119 GCACATTTATCAGGACACCCCAACCAACCAAGTGTAGTAAATAGCATGAGACCCGTG 60
XX DB |||||
XX 121 TTGTGTTTAAATTAATT 137
XX QY |||||
XX 59 TTGTGTTTAAATTAATT 43
XX DB |||||
XX
XX RESULT 4
XX ADD32504/c
XX ID ADD32504 standard; DNA; 455 BP.
XX

```

```

AC ADD32504;
XX
XX 15-JAN-2004 (first entry)
XX
XX Human mitochondrial DNA sequence SEQ ID NO:274.
XX
XX ds; human; array; mitochondrial; hybridisation; energy-metabolism;
XX KW mitochondrial disease; oxidative phosphorylation dysfunction;
XX KW oxidative stress; apoptosis; aging.
XX
XX Homo sapiens.
XX
XX WO2003020220-A2.
XX
XX 13-MAR-2003.
XX
XX 30-AUG-2002; 2002WO-US027886.
XX
XX 30-AUG-2001; 2001US-0316232P.
XX PR
XX 31-AUG-2001; 2001CA-02356540.
XX
XX (UYEM-) UNIV EMORY.
XX PA
XX Wallace DC, Levy S, Kerstann K, Procaccio V;
XX PI
XX WPI; 2003-300821/29.
XX
XX Array containing probes for genes involved in mitochondrial biology,
XX PT useful for determining mitochondrial biology gene expression profiles for
XX PT use in diagnosing pathologies and identifying biochemical pathways.
XX
XX Claim 2; SEQ ID NO 274; 201pp; English.
XX
XX The invention relates to a novel array comprising at least two isolated
XX CC nucleotide molecules, each molecule having a sequence capable of uniquely
XX CC hybridising to a nucleic acid molecule which is an expression product of
XX CC a gene involved in mitochondrial biology. The array comprises two or more
XX CC isolated nucleic acid molecules or spots, each molecule having a sequence
XX CC chosen from sequence of 394 human probes and 2046 mouse probes. An array
XX CC of the invention is useful for determining an expression profile of a
XX CC mouse or human sample containing nucleic acid, by contacting the array
XX CC with the sample under conditions allowing selective hybridisation, and
XX CC measuring hybridisation of nucleic acid in the sample to the array to
XX CC produce an expression profile. The array is also useful for determining
XX CC an expression profile of a first labelled sample containing nucleic acid
XX CC relative to a second, differentially labelled sample containing nucleic
XX CC acid. The second sample is a reference or a standard. An array is useful
XX CC for determining an expression profile diagnostic of an energy-metabolism-
XX CC related physiological condition. An array of the invention is useful for
XX CC determining mitochondrial biology gene expression profiles of organisms,
XX CC such as human, mice and closely related species, tissue and organs of
XX CC such organisms, which are useful for determining expression profiles
XX CC diagnostic of energy metabolism-related physiological conditions
XX CC diagnosing such physiological conditions, identifying biochemical
XX CC pathways, genes, and mutations involved in such physiological conditions,
XX CC identifying therapeutic agents useful for preventing and/or treating such
XX CC physiological conditions, evaluating and/or monitoring the efficacy of
XX CC such therapies, and creating and identifying animal models of human
XX CC energy metabolism-related physiological conditions. An array is also
XX CC useful for defining expression signatures or profiles for mitochondrial
XX CC diseases, as well as distinguishing clinical disorders that result from
XX CC oxidative phosphorylation (OXPHOS) dysfunction, oxidative stress,
XX CC apoptosis and aging. An array of the invention contains probes of genes
XX CC not previously recognised to participate in mitochondrial biology. The
XX CC sequences shown in ADD32231-ADD33223 represent human mitochondrial DNA
XX CC clones used to make the probes of the invention. Some sequences are not
XX CC present, these are SEQ ID NO's 295, 1174, 1213, 1700, 1728, 1730, 1905,
XX CC 1906, 2408 and 2643.
XX
XX Sequence 455 BP; 127 A; 80 C; 63 G; 185 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 137; DB 9; Length 455;
XX Best Local Similarity 100.0%; Pred. No. 2.3e-33;

```

Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATAAGGAAAAAGAAATAACGCAATGACAAAGTGTGCAAGCTGGAAGTCAAGTGT 60  
 DB 155 ATTATAAGGAAAAAGAAATAACGCAATGACAAAGTGTGCAAGCTGGAAGTCAAGTGT 96

QY 61 GCACAAATTATCAGGAGAACACCCCAACCAACCAAGTGTAGGAGTAAATAGCATGAGAGCCGTG 120  
 DB 95 GCACAAATTATCAGGAGAACACCCCAACCAACCAAGTGTAGGAGTAAATAGCATGAGAGCCGTG 36

QY 121 TTGTGATTTAATTAATT 137  
 DB 35 TTGTGATTTAATTAATT 19

RESULT 5  
 ABK86497  
 ID ABK86497 standard; cDNA; 996 BP.  
 AC ABK86497;  
 XX  
 XX  
 DT 27-AUG-2002 (first entry)  
 XX  
 DE Human Apo-dystrophin-4 cDNA.  
 XX  
 KW Human; ss; gene; apo-dystrophin-4; inversion sequence; gene therapy;  
 protein truncation; muscular dystrophy; leukaemia.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..996  
 FT /product= "Apo-dystrophin-4 protein appearing as  
 AAU98739"  
 FT /partial  
 FT /note= "No start or stop codon shown"  
 FT /transl\_except= (pos:7..9,aa:Xaa)  
 FT /transl\_except= (pos:49..51,aa:Xaa)  
 FT /transl\_except= (pos:187..189,aa:Xaa)  
 FT /transl\_except= (pos:202..204,aa:Xaa)  
 FT /transl\_except= (pos:217..219,aa:Xaa)  
 FT /transl\_except= (pos:253..255,aa:Xaa)  
 FT /transl\_except= (pos:265..270,aa:Xaa-Xaa)  
 FT /transl\_except= (pos:394..396,aa:Xaa)  
 FT /transl\_except= (pos:547..549,aa:Xaa)  
 FT /transl\_except= (pos:565..567,aa:Xaa)  
 FT /transl\_except= (pos:616..618,aa:Xaa)  
 FT /transl\_except= (pos:649..651,aa:Xaa)  
 FT /transl\_except= (pos:682..684,aa:Xaa)  
 FT /transl\_except= (pos:709..711,aa:Xaa)  
 FT /transl\_except= (pos:790..792,aa:Xaa)  
 FT /transl\_except= (pos:841..843,aa:Xaa)  
 FT /transl\_except= (pos:850..852,aa:Xaa)  
 FT /transl\_except= (pos:880..882,aa:Xaa)  
 FT /transl\_except= (pos:907..909,aa:Xaa)  
 FT /transl\_except= (pos:952..954,aa:Xaa)  
 FT /transl\_except= (pos:982..984,aa:Xaa)  
 FT /transl\_except= (pos:988..990,aa:Xaa)  
 FT /note= "Xaa= unknown, encoded by in frame stop codon"

XX  
 PN GB2368064-A.  
 XX  
 PD 24-APR-2002.  
 XX  
 PF 16-JAN-2001; 2001GB-00001124.  
 XX  
 PR 30-SEP-2000; 2000US-0237079P.  
 XX  
 PA (IMCR ) IMPERIAL CANCER RES TECHNOLOGY LTD.  
 PA (BARB/) BARBER E.  
 XX  
 PI Barber E;

XX WPI; 2002-429042/46.  
 DR P-PSDB; AAU98739.  
 XX  
 PT New human regulatory polynucleotide, useful for treating disorders  
 associated with protein truncation, particularly muscular dystrophy, and  
 related peptides and antibodies.  
 PT  
 XX  
 PS Claim 5; Page 170-172; 222pp; English.  
 XX  
 CC The invention relates to a polynucleotide (I) comprising, or consisting  
 of, apo-dystrophin-4 inversion sequence appearing as ABK86496, or its  
 functional equivalents (e.g. the apo-dystrophin-4 cDNA sequence appearing  
 as ABK86497). Also included are polynucleotides that hybridize to either  
 strand of (I), a vector containing (I), a cell containing (I) or the  
 vector, proteins and peptides encoded by (I), a protein homologous with  
 human dystrophin that is expressed on cell surfaces in vivo antibodies  
 (Ab) specific for the protein and method of screening for leukemia cells  
 by analysing DNA for presence of (I) or by detecting presence of (II).  
 CC The apo-dystrophin-4 inversion sequence is a regulatory element that  
 controls expression (transcription and translation) of associated DNA,  
 and may allow read-through of stop codons. The apo-dystrophin-4 inversion  
 sequence is used in gene therapy of diseases associated with truncation  
 of proteins, particularly muscular dystrophy and also leukaemia, but more  
 generally (I) is a regulatory sequence used to control expression of any  
 attached gene. Analysis of DNA for (I), or detection of proteins (II)  
 encoded by (I), can be used to screen for leukaemic cells and related  
 diseases. Antibodies raised against (II) can be used therapeutically, to  
 inhibit (II) activity, also to detect (II) in screening assays. The  
 present sequence is the cDNA sequence for human apo-dystrophin-4  
 containing a plurality of stop codons some of which may be read through  
 due to the presence of (I) in the apo-dystrophin-4 gene

XX  
 SQ Sequence 996 BP; 334 A; 152 C; 195 G; 315 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 137; DB 6; Length 996;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-33;  
 Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATAAGGAAAAAGAAATAACGCAATGACAAAGTGTGCAAGCTGGAAGTCAAGTGT 60  
 DB 860 ATTATAAGGAAAAAGAAATAACGCAATGACAAAGTGTGCAAGCTGGAAGTCAAGTGT 919

QY 61 GCACAAATTATCAGGAGAACACCCCAACCAACCAAGTGTAGGAGTAAATAGCATGAGAGCCGTG 120  
 DB 920 GCACAAATTATCAGGAGAACACCCCAACCAACCAAGTGTAGGAGTAAATAGCATGAGAGCCGTG 979

QY 121 TTGTGATTTAATTAATT 137  
 DB 980 TTGTGATTTAATTAATT 996

RESULT 6  
 ABK86462  
 ID ABK86462 standard; cDNA; 1230 BP.  
 AC ABK86462;  
 XX  
 XX 27-AUG-2002 (first entry)  
 DT  
 XX Human Apo-dystrophin-4 full length cDNA.  
 DE  
 KW Human; ss; gene; apo-dystrophin-4; inversion sequence; gene therapy;  
 protein truncation; muscular dystrophy; leukaemia.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..1230  
 FT /\*tag= a  
 FT /product= "Apo-dystrophin-4"  
 FT /partial  
 FT /note= "No start or stop codon shown"



FT FT /transl\_except= (pos:142..144,aa:Xaa)  
 FT FT /transl\_except= (pos:241..244,aa:Xaa)  
 FT FT /transl\_except= (pos:282..284,aa:Xaa)  
 FT FT /transl\_except= (pos:421..423,aa:Xaa)  
 FT FT /transl\_except= (pos:436..438,aa:Xaa)  
 FT FT /transl\_except= (pos:451..453,aa:Xaa)  
 FT FT /transl\_except= (pos:487..489,aa:Xaa)  
 FT FT /transl\_except= (pos:499..504,aa:Xaa-Xaa)  
 FT FT /transl\_except= (pos:628..630,aa:Xaa)  
 FT FT /transl\_except= (pos:781..783,aa:Xaa)  
 FT FT /transl\_except= (pos:799..801,aa:Xaa)  
 FT FT /transl\_except= (pos:850..852,aa:Xaa)  
 FT FT /transl\_except= (pos:883..885,aa:Xaa)  
 FT FT /transl\_except= (pos:916..920,aa:Xaa)  
 FT FT /transl\_except= (pos:945..947,aa:Xaa)  
 FT FT /transl\_except= (pos:1024..1026,aa:Xaa)  
 FT FT /transl\_except= (pos:1075..1077,aa:Xaa)  
 FT FT /transl\_except= (pos:1084..1086,aa:Xaa)  
 FT FT /transl\_except= (pos:1114..1116,aa:Xaa)  
 FT FT /transl\_except= (pos:1141..1143,aa:Xaa)  
 FT FT /transl\_except= (pos:1186..1188,aa:Xaa)  
 FT FT /transl\_except= (pos:1216..1218,aa:Xaa)  
 FT FT /transl\_except= (pos:1222..1224,aa:Xaa)  
 FT FT /note= "Xaa= unknown, encoded by in frame stop codon"

GB2368064-A.

XX XX 24-APR-2002.

XX XX 16-JAN-2001; 2001GB-00001124.

XX XX 30-SEP-2000; 2000US-0237079P.

XX XX (IMCR ) IMPERIAL CANCER RES TECHNOLOGY LTD.

XX XX (BARB/) BARBER E.

XX XX Barber E;

XX XX WPI; 2002-429042/46.

XX XX P-PSDB; AAU98729.

XX FT New human regulatory polynucleotide, useful for treating disorders  
 XX FT associated with protein truncation, particularly muscular dystrophy, and  
 XX FT related peptides and antibodies.

XX PS Disclosure; Fig 4; 222pp; English.

XX CC The invention relates to a polynucleotide (I) comprising, or consisting  
 CC of, apo-dystrophin-4 inversion sequence appearing as ABK86496, or its  
 CC functional equivalents (e.g. the apo-dystrophin-4 cDNA sequence appearing  
 CC as ABK86497). Also included are polynucleotides that hybridize to either  
 CC strand of (I), a vector containing (I), a cell containing (I) or the  
 CC vector, proteins and peptides encoded by (I), a protein homologous with  
 CC human dystrophin that is expressed on cell surfaces in vivo antibodies  
 CC (Ab) specific for the protein and method of screening for leukemia cells  
 CC by analysing DNA for presence of (I) or by detecting presence of (II).  
 CC The apo-dystrophin-4 inversion sequence is a regulatory element that  
 CC controls expression (transcription and translation) of associated DNA,  
 CC and may allow read-through of stop codons. The apo-dystrophin-4 inversion  
 CC sequence is used in gene therapy of diseases associated with truncation  
 CC of proteins, particularly muscular dystrophy and also leukaemia, but more  
 CC generally (I) is a regulatory sequence used to control expression of any  
 CC attached gene. Analysis of DNA for (I), or detection of proteins (II)  
 CC encoded by (II), can be used to screen for leukaemic cells and related  
 CC diseases. Antibodies raised against (II) can be used therapeutically, to  
 CC inhibit (II) activity, also to detect (II) in screening assays. The  
 CC present sequence is the full length cDNA sequence for human apo-  
 CC dystrophin-4 containing a plurality of stop codons some of which may be  
 CC read through due to the presence of (I) in the apo-dystrophin-4 gene

XX SQ Sequence 1230 BP; 404 A; 189 C; 259 G; 378 T; 0 U; 0 Other;

Query Match

100.0%; Score 137; DB 6; Length 1230;

Best Local Similarity 100.0%; Pred. No. 3e-33;  
 Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATTATTAAGGAAAGAAAGAAATACGCAATGCGCAAGTGGTGAAGCTGTGAACCTCAGGTGT 60  
 Db 1094 ATTATTAAGGAAAGAAAGAAATACGCAATGCGCAAGTGGTGAAGCTGTGAACCTCAGGTGT 1153  
 QY 61 GCACAATTATCAGGAACACCCCAAAACCAAAAGTGAAGTGAATATGATGAGAGCCGTG 120  
 Db 1154 GCACAATTATCAGGAACACCCCAAAACCAAAAGTGAAGTGAATATGATGAGAGCCGTG 1213  
 QY 121 TTTCATGTTTAATTAATT 137  
 Db 1214 TTTCATGTTTAATTAATT 1230  
 RESULT 7  
 ID ABK86463  
 XX ABK86463 standard; cDNA; 1234 BP.  
 AC ABK86463;  
 XX 27-AUG-2002 (first entry)  
 DT Human Apo-dystrophin-4 full cDNA fragment #1.  
 DE  
 XX Human; ss; apo-dystrophin-4; inversion sequence; gene therapy;  
 KW protein truncation; muscular dystrophy; leukaemia.  
 XX Homo sapiens.  
 XX GB2368064-A.  
 XX 24-APR-2002.  
 XX 16-JAN-2001; 2001GB-00001124.  
 XX 30-SEP-2000; 2000US-0237079P.  
 XX (IMCR ) IMPERIAL CANCER RES TECHNOLOGY LTD.  
 XX (BARB/) BARBER E.  
 XX Barber E;  
 XX WPI; 2002-429042/46.  
 XX New human regulatory polynucleotide, useful for treating disorders  
 XX associated with protein truncation, particularly muscular dystrophy, and  
 XX related peptides and antibodies.  
 XX Disclosure; Fig 9; 222pp; English.  
 CC The invention relates to a polynucleotide (I) comprising, or consisting  
 CC of, apo-dystrophin-4 inversion sequence appearing as ABK86496, or its  
 CC functional equivalents (e.g. the apo-dystrophin-4 cDNA sequence appearing  
 CC as ABK86497). Also included are polynucleotides that hybridize to either  
 CC strand of (I), a vector containing (I), a cell containing (I) or the  
 CC vector, proteins and peptides encoded by (I), a protein homologous with  
 CC human dystrophin that is expressed on cell surfaces in vivo antibodies  
 CC (Ab) specific for the protein and method of screening for leukemia cells  
 CC by analysing DNA for presence of (I) or by detecting presence of (II).  
 CC The apo-dystrophin-4 inversion sequence is a regulatory element that  
 CC controls expression (transcription and translation) of associated DNA,  
 CC and may allow read-through of stop codons. The apo-dystrophin-4 inversion  
 CC sequence is used in gene therapy of diseases associated with truncation  
 CC of proteins, particularly muscular dystrophy and also leukaemia, but more  
 CC generally (I) is a regulatory sequence used to control expression of any  
 CC attached gene. Analysis of DNA for (I), or detection of proteins (II)  
 CC encoded by (II), can be used to screen for leukaemic cells and related  
 CC diseases. Antibodies raised against (II) can be used therapeutically, to  
 CC inhibit (II) activity, also to detect (II) in screening assays. The  
 CC present sequence is the full length cDNA sequence for human apo-  
 CC dystrophin-4 containing a plurality of stop codons some of which may be  
 CC read through due to the presence of (I) in the apo-dystrophin-4 gene  
 CC of mouse genomic DNA in the Apo-dystrophin-4 gene region

QY	1	ATTATAAGGAAAGAAATACCGAATGGACAAAGTGTGAAGCTGTGAACCTCAGGTGT	60
Db	2033	ATTATAAGGAAAGAAATACCGAATGGACAAAGTGTGAAGCTGTGAACCTCAGGTGT	1974
QY	61	GCACAAATTATCAGGAACACACCCCAAAACCCAAAGTGAGGTAGAAATAGCATGAGAACCCGTG	120
Db	1973	GCACAAATTATCAGGAACACACCCCAAAACCCAAAGTGAGGTAGAAATAGCATGAGAACCCGTG	1914
QY	121	TTTGATGTTAANTTAATT	137
Db	1913	TTTGATGTTAANTTAATT	1897
RESULT 9			
ABK82000/c			
ID	ABK82000	standard; DNA; 8689 BP.	
XX	ABK82000;		
AC	XX		
DT	XX		
DE	13-AUG-2002	(first entry)	
DE	XX	DNA encoding mini-dystrophin protein deltaH2-R19.	
KW	XX	Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;	
KW	XX	Duchenne's muscular dystrophy; DMD; dystrophin; ds.	
OS	XX	Homo sapiens.	
OS	XX	Synthetic.	
XX	FN	WC0200229056-A2.	
XX	XX		
PD	11-APR-2002.		
XX	XX		
PF	04-OCT-2001; 2001WO-US031126.		
XX	XX		
PR	06-OCT-2000; 2000US-02398849P.		
XX	XX		
PA	(UNMI ) UNIV MICHIGAN.		
XX	XX	Chamberlain JS, Harper SQ;	
PI	XX	WPI; 2002-435334/46.	
DR	XX		
XX	XX		
PT	XX	A composition for preparing therapeutic drugs, has a mini-dystrophin	
PT	PT	peptide comprising a specific number of spectrin-like repeat domains, or	
PT	PT	a nucleic acid sequence encoding the mini-dystrophin peptide.	
XX	XX		
PS	XX	Disclosure; Fig 15; 145pp; English.	
XX	XX		
CC	XX	The invention describes a composition comprising a mini-dystrophin	
CC	CC	peptide comprising a spectrin-like repeat domain, where the domain	
CC	CC	comprises n spectrin-like repeats, and contains no more than n spectrin-	
CC	CC	like repeats, where n is an even number between 4-24, or a nucleic acid	
CC	CC	encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the	
CC	CC	polynucleotide encoding it is useful as a medicament, for preparing a	
CC	CC	drug for therapeutic application and in the preparation of a composition	
CC	CC	for treatment of muscle disease, e.g. Duchenne's muscular dystrophy	
CC	CC	(DMD). This sequence represents a mini-dystrophin sequence of the	
CC	CC	invention	
XX	XX		
SQ	XX	Sequence 8689 BP; 2721 A; 1804 C; 1861 G; 2303 T; 0 U; 0 Other;	
Query Match		100.0%; Score 137; DB 6; Length 8689;	
Best Local Similarity		100.0%; Pred. No. 5.2e-33;	
Matches 137; Conservative		0; Mismatches 0; Indels 0; Gaps 0	
QY	1	ATTATAAGGAAAGAAATACCGAATGGACAAAGTGTGAAGCTGTGAACCTCAGGTGT	60
Db	8031	ATTATAAGGAAAGAAATACCGAATGGACAAAGTGTGAAGCTGTGAACCTCAGGTGT	7977
QY	61	GCACAAATTATCAGGAACACCCCAAAACCCAAAGTGAGGTAGAAATAGCATGAGAACCCGTG	120
Db	7971	GCACAAATTATCAGGAACACCCCAAAACCCAAAGTGAGGTAGAAATAGCATGAGAACCCGTG	7911

QY	121	TTTGATGTTAAATTAAT	137
DB	7911	TTTGATGTTAAATTAAT	7895
RESULT 10			
ID	ABK82002/c	ABK82002 standard; DNA; 11443 BP.	
AC	ABK82002;		
XX			
DT	13-AUG-2002	(first entry)	
DE		DNA encoding mini-dystrophin protein deltaR9-R16.	
XX			
KW		Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;	
KW		Duchenne's muscular dystrophy; DMD; dystrophin; ds.	
XX			
OS		Homo sapiens.	
OS		Synthetic.	
XX			
PN		WO200229056-A2.	
XX			
PD		11-APR-2002.	
XX			
PF		04-OCT-2001; 2001WO-US031126.	
XX			
PR		06-OCT-2000; 2000US-0238848P.	
XX			
PA		(UNMI ) UNIV MICHIGAN.	
XX			
PI		Chamberlain JS, Harper SQ;	
XX			
DR		WPI; 2002-435334/46.	
XX			
PT		A composition for preparing therapeutic drugs, has a mini-dystrophin	
PT		peptide comprising a specific number of spectrin-like repeat domains, or	
PT		a nucleic acid sequence encoding the mini-dystrophin peptide.	
XX			
PS		Disclosure; Fig 17; 145pp; English.	
XX			
CC		The invention describes a composition comprising a mini-dystrophin	
CC		peptide comprising a spectrin-like repeat domain, where the domain	
CC		comprises n spectrin-like repeats, and contains no more than n spectrin-	
CC		like repeats, where n is an even number between 4-24, or a nucleic acid	
CC		encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the	
CC		polynucleotide encoding it is useful as a medicament, for preparing a	
CC		drug for therapeutic application and in the preparation of a composition	
CC		for treatment of muscle disease, e.g. Duchenne's muscular dystrophy	
CC		(DMD). This sequence represents a mini-dystrophin sequence of the	
CC		invention	
XX			
SQ		Sequence 11443 BP; 3707 A; 2339 C; 2502 G; 2895 T; 0 U; 0 Other;	
		Query Match 100.0%; Score 137; DB 6; Length 11443;	
		Best Local Similarity 100.0%; Pred. No. 5.6e-33;	
		Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	ATTATAAGGAAAGAAATTAACGCAATGGACAAAGTGTGAAGCTGTGAAGCTGTGAGTGT	60
DB	10785	ATTATAAGGAAAGAAATTAACGCAATGGACAAAGTGTGAAGCTGTGAAGCTGTGAGTGT	10726
QY	61	GCACAAATTATCAGGAACACCCCAAAACCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGT	120
DB	10725	GCACAAATTATCAGGAACACCCCAAAACCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGT	10666
QY	121	TTTGATGTTAAATTAAT	137
DB	10665	TTTGATGTTAAATTAAT	10649
RESULT 11			
ID	ABK81959	ABK81959 standard; DNA; 13957 BP.	
AC	ABK81959;		
XX			
DT	13-AUG-2002	(first entry)	
DE		cDNA encoding human dystrophin.	
XX			
KW		Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;	
KW		Duchenne's muscular dystrophy; DMD; dystrophin; human; gene; ds.	
XX			
OS		Homo sapiens.	
XX			
PN		WO200229056-A2.	
XX			
PD		11-APR-2002.	
XX			
PF		04-OCT-2001; 2001WO-US031126.	
XX			
PR		06-OCT-2000; 2000US-0238848P.	
XX			
PA		(UNMI ) UNIV MICHIGAN.	
XX			
PI		Chamberlain JS, Harper SQ;	
XX			
DR		WPI; 2002-435334/46.	
XX			
PT		A composition for preparing therapeutic drugs, has a mini-dystrophin	
PT		peptide comprising a specific number of spectrin-like repeat domains, or	
PT		a nucleic acid sequence encoding the mini-dystrophin peptide.	
XX			
PS		Example 2; Fig 1; 145pp; English.	
XX			
CC		The invention describes a composition comprising a mini-dystrophin	
CC		peptide comprising a spectrin-like repeat domain, where the domain	
CC		comprises n spectrin-like repeats, and contains no more than n spectrin-	
CC		like repeats, where n is an even number between 4-24, or a nucleic acid	
CC		encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the	
CC		polynucleotide encoding it is useful as a medicament, for preparing a	
CC		drug for therapeutic application and in the preparation of a composition	
CC		for treatment of muscle disease, e.g. Duchenne's muscular dystrophy	
CC		(DMD). This sequence represents a mini-dystrophin sequence of the	
CC		invention	
XX			
SQ		Sequence 13957 BP; 4602 A; 2781 C; 3122 G; 3452 T; 0 U; 0 Other;	
		Query Match 100.0%; Score 137; DB 6; Length 13957;	
		Best Local Similarity 100.0%; Pred. No. 6e-33; 0; Indels 0; Gaps 0;	
		Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	ATTATAAGGAAAGAAATTAACGCAATGGACAAAGTGTGAAGCTGTGAAGCTGTGAGTGT	60
DB	13299	ATTATAAGGAAAGAAATTAACGCAATGGACAAAGTGTGAAGCTGTGAGTGTGAGTGT	13240
QY	61	GCACAAATTATCAGGAACACCCCAAAACCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGT	120
DB	13239	GCACAAATTATCAGGAACACCCCAAAACCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGT	13180
QY	121	TTTGATGTTAAATTAAT	137</



KW adenosine deaminase deficiency; severe combined immune deficiency; PAH;  
KW beta-chain; haemoglobin gene; beta-thalassaemia; sickle cell disease;  
KW low density lipoprotein gene; familial hypercholesterolaemia;  
KW hypoxanthine-guanine phosphoribosyltransferase; Lesch-Nyhan syndrome;  
KW phenylalanine hydroxylase gene; gene therapy; phenylketonuria;  
KW dystrophin gene; muscular dystrophy; cystic fibrosis; immunostimulant;  
KW human cystic fibrosis transmembrane conductance regulator gene;  
KW antianemic; antilipemic; noctropic; cytosatic; dermatological; human;  
KW alpha-1-antitrypsin; lysosomal glucocerebrosidase; ADA; HPRT;  
KW lysosomal arylsulfatase A; omithine transcarbamylase; ARSA; OTC; NP;  
KW purin nucleoside phosphorylase; gene; ds.  
OS Homo sapiens.  
XX US2002102731-A1.  
XX 01-AUG-2002.  
XX 12-FEB-2001; 2001US-00782378.  
XX 02-OCT-2000; 2000US-0237747P.  
XX (UINY ) UNIV NEW YORK STATE RES FOUND.  
XX PA Hearing P, Bahou WF, Sandalon Z, Gnatenko DV;  
XX PI WPI; 2002-690619/74.  
XX DR  
XX PT Producing vector, by introducing vector having nucleotide sequence.  
XX PT adeno-virus inverted terminal repeats and packaging sequence, and adeno-  
XX PT associated virus terminal repeat, into cell, and culturing cell.  
XX PS Disclosure; Page 122-128; 191pp; English.  
XX CC The present invention relates to a new method of producing a vector. The  
XX CC method involves introducing recombinant vector having nucleotide sequence  
XX CC (NS) having 5' and 3' end, left and right inverted terminal repeats of  
XX CC adeno-virus flanking NS, adeno-virus packaging sequence linked to inverted  
XX CC terminal repeat, and adeno-associated virus terminal repeat linked to 3'  
XX CC end of NS, into cell expressing adeno-virus early gene lacking from vector  
XX CC ; and culturing cell to produce another vector. The method is useful for  
XX CC generating vectors, especially M4D vectors. The method is useful in  
XX CC transferring nucleotide sequences of interest into a cell, for gene  
XX CC transfer applications (e.g. gene therapy) in vitro, ex vivo and in vivo.  
XX CC The nucleotide sequences are useful for treating diseases associated with  
XX CC it, i.e. adenosine deaminase gene associated with adenosine deaminase  
XX CC deficiency with severe combined immune deficiency, beta-chain of  
XX CC haemoglobin gene associated with beta-thalassaemia and sickle cell  
XX CC disease, receptor for low density lipoprotein gene associated with  
XX CC familial hypercholesterolaemia, hypoxanthine-guanine  
XX CC phosphoribosyltransferase associated with Lesch-Nyhan syndrome,  
XX CC phenylalanine hydroxylase (PAH) gene associated with phenylketonuria,  
XX CC dystrophin gene associated with muscular dystrophy, and human cystic  
XX CC fibrosis transmembrane conductance regulator gene associated with cystic  
XX CC fibrosis. The present nucleic acid sequence represents a human disease  
XX CC gene sequence that was used in the methods of the invention  
XX SQ Sequence 13957 BP; 4602 A; 2781 C; 3122 G; 3452 T; 0 U; 0 Other;  
Query Match 100.0%; Score 137; DB 6; Length 13957;  
Best Local Similarity 100.0%; Pred. No. 6e-33;  
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATTATAAGGAAAAGAAAATACGCAATGGCAAGTGGTGAAGCTGTGAACCTCAGGTGT 60  
DB 13299 ATTATAAGGAAAAGAAAATACGCAATGGCAAGTGGTGAAGCTGTGAACCTCAGGTGT 13240  
QY 61 GCACATTATCAGGAACACCCCAACCAACCAAGTAGAGTGAATAGCATGAGAGCCGTTG 120  
DB 13239 GCACATTATCAGGAACACCCCAACCAACCAAGTAGAGTGAATAGCATGAGAGCCGTTG 13180  
QY 121 TTGTGTTTAATTAATT 137  
|||||

DB 13179 TTGTGTTTAATTAATT 13163  
RESULT 15  
ABS70403/C  
ID ABS70403 standard; cDNA; 13977 BP.  
XX AC ABS70403;  
XX AC ABS70403;  
XX DT 27-NOV-2002 (first entry)  
XX DE Human bone remodelling gene #60.  
XX KW Bone remodelling; osteoporosis; human; gene; ss.  
XX OS Homo sapiens.  
XX FN US6426186-B1.  
XX XX 30-JUL-2002.  
XX XX 18-JAN-2000; 2000US-00484970.  
XX XX 18-JAN-2000; 2000US-00484970.  
XX XX (INCY-) INCYTE GENOMICS INC.  
XX PI Jones KA, Volkmut W, Walker MG;  
XX DR WPI; 2002-673014/72.  
XX PT A combination of polynucleotides which are co-expressed with genes known  
XX PT to be involved in bone remodeling and osteoporosis are useful in an array  
XX PT for the diagnosis of bone remodeling and osteoporosis associated  
XX PT disorders.  
XX PS Claim 1; Col 185-198; 206pp; English.  
XX CC The invention relates to a combination comprising a number of  
XX CC substantially purified and isolated polynucleotides which are co-  
XX CC expressed with genes known to be involved in bone remodeling and  
XX CC osteoporosis. The invention is used to diagnose disorders associated with  
XX CC bone remodeling or osteoporosis. ABS70344-ABS70512 represent human bone  
XX CC remodelling genes of the invention  
XX SQ Sequence 13977 BP; 4596 A; 2765 C; 3120 G; 3453 T; 0 U; 43 Other;  
Query Match 92.0%; Score 126; DB 6; Length 13977;  
Best Local Similarity 99.3%; Pred. No. 1.8e-29;  
Matches 137; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
QY 1 ATTATAAGGAAAAGAAAATACGCAATGGCAAGTGGTGAAGCTGTGAACCTCAGGTGT 60  
DB 13306 ATTATAAGGAAAAGAAAATACGCAATGGCAAGTGGTGAAGCTGTGAACCTCAGGTGT 13247  
QY 61 GCACATTATCAGGAACACCCCAACCAACCAAGTAGAGTGAATAGCATGAGAGCCGTT 119  
DB 13246 GCACATTATCAGGAACACCCCAACCAACCAAGTAGAGTGAATAGCATGAGAGCCGTT 13187  
QY 120 GTTTGATGTTAATTAATT 137  
DB 13186 GTTTGATGTTAATTAATT 13169  
Search completed: April 6, 2004, 11:42:37  
Job time : 68.812 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: April 6, 2004, 03:22:40 ; Search time 574.36 Seconds  
(without alignments)  
10338.454 Million cell updates/sec

Title: US-09-966-264D-1  
Perfect score: 137  
Sequence: 1 attataaggaagaaagaaaa.....gtgttgatgtaattaatt 137

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21571516995 residues  
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.cm.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.ox.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*  
29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.tod.\*  
36: em.htg.nam.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	137	100.0	2148	11	G15848
C 2	137	100.0	2563	9	AK129855
C 3	137	100.0	2691	6	AX538618
C 4	137	100.0	4658	9	BC028720
C 5	137	100.0	8689	6	AX538622
C 6	137	100.0	11443	6	AX538624
C 7	137	100.0	13957	6	AX409637
C 8	137	100.0	13957	6	AX538581
C 9	137	100.0	13957	9	HUMDYS
C 10	137	100.0	98056	9	AC006061
C 11	126	92.0	13977	6	AR220819
C 12	102.8	75.0	13887	4	AF070485
C 13	100.6	73.4	13815	6	AX306153
C 14	100.6	73.4	13815	6	AX538582
C 15	100.6	73.4	13815	10	MUSDYS
C 16	100.6	73.4	13907	6	AR093392
C 17	100.6	73.4	19307	6	AR142592
C 18	100.6	73.4	189131	10	AL645477
C 19	100.6	73.4	272578	2	AC108338
C 20	100.6	73.4	279539	2	AC114184
C 21	68.4	49.9	13575	5	GGDYS
C 22	65	47.4	212120	2	AC023414
C 23	39.8	29.1	172734	10	AL528772
C 24	35.8	26.1	94665	3	AC099765
C 25	35.2	25.7	142000	9	AP005404
C 26	35.2	25.7	153228	2	AC139129
C 27	35.2	25.7	155949	9	AC091022
C 28	35.2	25.7	158663	2	AC069249
C 29	35.2	25.7	158682	2	AC023083
C 30	35.2	25.7	183365	2	AC015666
C 31	35.2	25.7	165789	10	AC111145
C 32	35.2	25.7	180573	9	AC027689
C 33	35	25.5	172422	2	BX649289
C 34	35	25.5	231578	2	BX548167
C 35	34.6	25.3	137678	8	AP005183
C 36	34.6	25.3	138653	8	AP005178
C 37	34.6	25.3	240279	2	AC126486
C 38	34.6	25.3	270456	2	AC111575
C 39	34.4	25.1	495	8	AY202075
C 40	34.4	25.1	50398	3	AC024771
C 41	34.4	25.1	106688	8	ATAC011436
C 42	34.4	25.1	206217	2	AC006754
C 43	34.2	25.0	112944	9	AC107082
C 44	34.2	25.0	169755	2	AP001199
C 45	34.2	25.0	171366	9	AC130303

ALIGNMENTS

RESULT 1  
G15848/c  
LOCUS  
DEFINITION human STS CHLC.UTR\_01924\_M18533. P56108 clone UTR\_01924\_M18533,  
sequence tagged site.  
ACCESSION G15848  
VERSION G15848.1 GI:1161737  
KEYWORDS STS; STS sequence; primer; sequence tagged site.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2148)  
AUTHORS Murray, J., Sheffield, V., Weber, J.L., Duyk, G. and Buetow, K.H.  
TITLE Cooperative Human Linkage Center

STANDARD  
G15848 2148 bp DNA linear STS 19-JAN-1996  
human STS CHLC.UTR\_01924\_M18533. P56108 clone UTR\_01924\_M18533,  
sequence tagged site.

Unpublished (1995)  
Synonyms: UTR\_01924\_M18533, CH1C.UTR\_01924\_M18533.T36152  
Contact: Dr. Jeffrey C. Murray  
UofI  
The University of Iowa  
Department of Pediatrics, Iowa City, IA 52242, USA  
Tel: (319) 356-3508  
Tel: (319) 356-3347  
Fax: (319) 356-3347  
Email: jeff-murray@uiowa.edu  
Primer A: AACGCATTTTGGTTGTTTA  
Primer B: GATATCAGCCCAAAAGGATG  
STS size: 189  
PCR Profile:

denature:	30 seconds at 94 degrees C
annealing:	75 seconds at 55 degrees C
extension:	15 seconds at 72 degrees C
PCR cycles:	27
extension:	6 minutes at 72 degrees C
Protocol:	
Template:	30ng genomic DNA
Primer:	each 1.5 pmole
dNTPs:	each 200 uM
Taq Polymerase:	0.3 units
Total Vol:	10 ul

Buffer:

MgCl <sub>2</sub> :	1.5mM
KCl:	50mM
Tris:	10mM
pH:	8.3

Prepared with primer pairs derived from M18533.

```

FEATURES
Location/Qualifiers
1..2148
/organism="Homo sapiens"
/mol_type="genomic"
/db_xref="taxon:9606"
350..538
STS
primer_bind
350..369
primer_bind
complement(519..538)
ORIGIN

```

	Query Match	100.0%	Score 137;	DB 11;	Length 2148;
	Best Local Similarity	100.0%;	Pred. No. 7.7e-28;		
	Matches 137;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ATTATAAGGAAAAGAAATTAACGCATCGACAAGTGCTCAAGCTCTGAACTCAGGTGT	60		
Db	2033	ATTATAAGGAAAAGAAATTAACGCATCGACAAGTGCTGAAGCTGTGAACTCAGGTGT	1974		
QY	61	GCACAAATTATCAGGAACACCCCAAAACCAAAAGTCAGGTAGAAAATGACATGAGAAGCCGCTG	120		
Db	1973	GCACAAATTATCAGGAACACCCCAAAACCAAAAGTCAGGTAGAAAATGACATGAGAAGCCGCTG	1914		
QY	121	TTTGTATCTTAATTAATT	137		
Db	1913	TTTGTATCTTAATTAATT	1897		

[illegible]

Hata, H., Nakagawa, K., Mizuno, S., Morinaga, M., Kawamura, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Nishikawa, T., Sugiyama, A., Kawakami, B., Nagai, K., Isogai, T. and Sugano, S.  
NEDO human cDNA sequencing project  
Unpublished  
2 (bases 1 to 2563)  
Sugano, S. and Suzuki, Y.  
Direct Submission  
Submitted (31-JUL-2003) Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure, Human Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan  
(E-mail: fldn@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)  
NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology (RAB); cDNA library construction and 5'-end one pass sequencing; Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure, Human Genome Center; 3'-end one pass sequencing; RAB; clone selection for full insert sequencing; RAB and Helix Research Institute.  
Location/Qualifiers  
1. .2563  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="HRT03668"  
/tissue type="heart"  
/clone\_lib="HRT"  
/notes="Cloning vector: pME189FL3"

ORIGIN

	Query Match	100.0%;	Score 137;	DB 9;	Length 2563;
	Best Local Similarity	100.0%;	Pred. No. 7.5e-28;		
	Matches 137;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	ATTATTAAGGAAAGAAAGAAATTAACGCAATCGACAAAGTGGTGAAGCTGTGAACTCAGGTGT	60		
Db	1905	ATTATTAAGGAAAGAAAGAAATTAACGCAATCGACAAAGTGGTGAAGCTGTGAACTCAGGTGT	1846		
Qy	61	GCACAAATTATCAGGAACACCCCAAAACCAAAAGTCAGGTAGAGAAATAGCATGAGAAGCCGGT	120		
Db	1845	GCACAAATTATCAGGAACACCCCAAAACCAAAAGTCAGGTAGAGAAATAGCATGAGAAGCCGGT	1786		
Qy	121	TTTGATGTTTAATTAAATT	137		
Db	1785	TTTGATGTTTAATTAAATT	1769		

RESULT 3					PAT 23-NOV-2002
AX538618/c					
LOCUS			DNA	linear	
DEFINITION	Sequence 38 from Patent WO0229056.	2691 bp			
ACCSSION	AX538618				
VERSION	AX538618.1	GI:25271161			
KEYWORDS	.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
REFERENCE	1				

```

REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1
Chamberlain,J.S. and Harper,S.Q.
Mini-dystrophin nucleic acid and peptide sequences
Patent: WO 0229056-A 38 11-APR-2002;
THE REGENTS OF THE UNIVERSITY OF MICHIGAN (US)
Location/Qualifiers
1. .2891
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

```

```

ORIGIN
Query Match 100.0%; Score 137; DB 6; Length 2691;
Best Local Similarity 100.0%; Pred. No. 7.5e-28;
/WD_Art1- taxon:2600

```

Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATAAGGAAAAAGAAATAACCAATGACAGTGGTGAAGTGTGAACCTCAGGTGT 60  
 Db 2033 ATTATAAGGAAAAAGAAATAACCAATGACAGTGGTGAAGTGTGAACCTCAGGTGT 1974

QY 61 GCACAATTATCAGGAACACCCCAACCAACCAAGTAGGAGTAGAAATAGCATGAGAACCCGTG 120  
 Db 1973 GCACAATTATCAGGAACACCCCAACCAACCAAGTAGGAGTAGAAATAGCATGAGAACCCGTG 1914

QY 121 TTGTGATGTTAATTAATT 137  
 Db 1913 TTGTGATGTTAATTAATT 1897

RESULT 4  
 BC028720/c  
 LOCUS  
 DEFINITION Homo sapiens dystrophin (muscular dystrophy, Duchenne and Becker types), transcript variant Dp71b, mRNA (cDNA clone IMAGE:4822807), complete cds.  
 ACCESSION BC028720  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM Homo sapiens (human)

REFERENCE  
 AUTHORS Strausberg R.D., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altshuler S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan K., Moore T., Max S.I., Wang J., Heideh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Frange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S.J., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahy J., Helton E., Kettman M., Madan A.C., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S., Krzywinski M.I., Skalska U., Smalley D.E., Schnerch A., Schein J.E., Jones S.J. and Marra M.A.  
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 22388257  
 12477932

TITLE  
 JOURNAL  
 MEDLINE  
 PubMed  
 REFERENCE  
 AUTHORS Strausberg R.  
 TITLE Direct Submission  
 JOURNAL Submitted (29-APR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK  
 COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgabbs-r@mail.nih.gov](mailto:cgabbs-r@mail.nih.gov)  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA sequencing by: Institute for Systems Biology  
<http://www.systemsbio.org>  
 Contact: [amad@systemsbiology.org](mailto:amad@systemsbiology.org)  
 Anup Madan, Jessica Fahey, Erin Helton, Mark Kettman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAX Plate: 46 Row: j Column: 3  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 5032296  
 This clone has the following problem: The cds is short compared to the longest cds in the locus.

FEATURES  
 source  
 Location/Qualifiers  
 1..4658  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4822807"  
 /tissue\_type="Testis"  
 /clone\_lib="NIH MGC 97"  
 /lab\_host="DH10E"  
 /note="Vector: pBluescript"

gene  
 1..4658  
 /gene="DMD"  
 /notes="synonyms: BMD, DXS142, DXS164, DXS206, DXS230, DXS233, DXS268, DXS269, DXS270, DXS272"  
 /db\_xref="LocusID:1756"  
 /db\_xref="MIM:300377"  
 130..2037  
 /codon\_start=1  
 /product="dystrophin, Dp71b isoform"  
 /protein\_id="AAH28720.1"  
 /db\_xref="GI:20379676"  
 /db\_xref="LocusID:1756"  
 /translation="MREQLKHETQTTWDPKMTLYQSLADLNVRFSAYRTAKML RFLKALCDLLSLSAACDALQNDQMDILOIINCLTTIYDRLEQEHNLVN VFLCDMCLNLNVYDTGTGRIIVLSFKTGIIISLCKAHLEDKVYLFQVASTGFL CDRRLGLLHDSIQIPLQGEVAVSGGNSVRSQFPANNKEIEMAAFLDPMR LEPQSWLPLFVRAAATAKQACNICKCPIIFGRYSKJKNFYDIDICQCFSG RVAKGHKYPWVEYCTPTTSGEDVDFKVLKFKRTKRYFAKHPRMGLYPQTVILE GNMETPVTILNFWPVDSPASSPOLSHDDTSHRIEYASRLAEMENSGSYLNDIS PNEIDDEHLIIOHYCSLNQDSPSPQSPQAPQILISLESEGERELIADLEENR NLCAEYDRLKQOEHKGLSPLSPMPMTSPQSPRDAELIAEAKLRCHKGLRLAEMR QILEDHNKQLESQHLRLQLLEQPOAEKNGTIVSSPSTSLQSRSSOPMLRLVVGSL QTSDSMGEEDLLSPDQDTGTGLEVNEQLNPFSSRGHNVGSLFEWADDLGRAMESL VSWMTDEEGAE"

misc\_feature  
 844..981  
 /notes="ZZ; Region: Zinc finger, ZZ type, zinc finger present in dystrophin, CBP/p300. ZZ in dystrophin binds calmodulin Putative zinc finger"  
 /db\_xref="CDD:pfam00569"

ORIGIN  
 Query Match 100.0%; Score 137; DB 9; Length 4658;  
 Best Local Similarity 100.0%; Pred. No. 7e-28;  
 Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATAAGGAAAAAGAAATAACCAATGACAGTGGTGAAGTGTGAACCTCAGGTGT 60  
 Db 3984 ATTATAAGGAAAAAGAAATAACCAATGACAGTGGTGAAGTGTGAACCTCAGGTGT 3925

QY 61 GCACAATTATCAGGAACACCCCAACCAACCAAGTAGGAGTAGAAATAGCATGAGAACCCGTG 120  
 Db 3924 GCACAATTATCAGGAACACCCCAACCAACCAAGTAGGAGTAGAAATAGCATGAGAACCCGTG 3865

QY 121 TTGTGATGTTAATTAATT 137  
 Db 3864 TTGTGATGTTAATTAATT 3848

RESULT 5  
 AX538622/c  
 LOCUS  
 DEFINITION Sequence 42 from Patent WO0229056.  
 ACCESSION AX538622  
 VERSION AX538622.1  
 KEYWORDS GI:25271171  
 SOURCE  
 ORGANISM synthetic construct  
 synthetic construct







```

RESULT 10
AC006061
DEFINITION Homo sapiens X BAC GSHB-19024 (Genome Systems Human BAC Library)
complete sequence.
ACCESSION AC006061
VERSION AC006061.1
KEYWORDS GI:4204246
SOURCE HTG.
ORGANISM Homo sapiens (human)
REFERENCE
AUTHORS Muzny,D., Arenson,A.D., Bouck,J., Brundage,E., Bunac,C., Chen,Z.,
Di,W., Ding,Y., Dugan,S., Durbin,J., Forcum,J., Garcia,C.,
Gorell,J.H., Gorrell,L.L., Hernandez,J., Jackson,L.,
Kondrjewski,N., Leal,B., Lichtarge,O., Liu,W., Logan,O., Lu,J.,
Martinez,C., Oswald,G., Pampell,L.R., Parish,B.J., Perez,L.,
Rashid,N.D., Rives,C., Scherer,S.E., Shen,H., Simon,M., Vo,Q.,
Williamson,A., Worley,K.C., Yu,W., Zhou,X., Nelson,D. and
Gibbs,R.A.
DIRECT SUBMISSION
Unpublished
2 (bases 1 to 98056)
Worley,K.C.
JOURNAL Submitted (26-NOV-1998) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (30-JAN-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 98056)
Worley,K.C.
DIRECT SUBMISSION
Unpublished
2 (bases 1 to 98056)
Worley,K.C.
JOURNAL Submitted (02-FEB-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 98056)
Worley,K.C.
DIRECT SUBMISSION
Unpublished
2 (bases 1 to 98056)
Worley,K.C.
JOURNAL Submitted (04-FEB-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
6 (bases 1 to 98056)
Worley,K.C.
DIRECT SUBMISSION
Unpublished
2 (bases 1 to 98056)
Worley,K.C.
JOURNAL Submitted (28-MAR-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
7 (bases 1 to 98056)
Worley,K.C.
DIRECT SUBMISSION
Unpublished
2 (bases 1 to 98056)
Worley,K.C.
JOURNAL Submitted (07-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
8 (bases 1 to 98056)
Worley,K.C.
DIRECT SUBMISSION
Unpublished
2 (bases 1 to 98056)
Worley,K.C.
JOURNAL Submitted (01-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT
INFORMATION: http://gc.bcm.tmc.edu:8088/home.html or email
gc-help@bcm.tmc.edu
CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.

```

Overlapping clones are noted at the beginning and end of the Features listing.

#### ANNOTATION OF FEATURES:

STSS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones, and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases.

FEATURES	source
Location/Qualifiers	1. 98056
Location/Qualifiers	/organism="Homo sapiens"
Location/Qualifiers	/mol_type="genomic DNA"
Location/Qualifiers	/db_xref="taxon:9606"
Location/Qualifiers	/chromosome="X"
Location/Qualifiers	/clone="GSHB-19024"
misc_feature	1. 1999
misc_feature	/notes="overlaps bases 151178..153177 of clone AC117405"
misc_feature	/functions="clone overlap"
repeat_region	40..82
repeat_region	/rpt_family="L2"
repeat_region	378..507
repeat_region	/rpt_family="L1MD3"
repeat_region	complement(518..570)
repeat_region	/rpt_family="(GGA)n"
repeat_region	571..729
repeat_region	/rpt_family="L1MC4"
repeat_region	730..987
repeat_region	/rpt_family="WSTD"
repeat_region	complement(931..1400)
repeat_region	/rpt_family="L2"
repeat_region	complement(1649..1779)
repeat_region	/rpt_family="WIR"
repeat_region	complement(1783..2212)
repeat_region	/rpt_family="L1R16A"
repeat_region	complement(3626..3656)
repeat_region	/rpt_family="(GA)n"
repeat_region	complement(4713..5074)
repeat_region	/rpt_family="THE1B"
repeat_region	5702..7108
repeat_region	/rpt_family="L2"
repeat_region	complement(8670..8734)
repeat_region	/rpt_family="WLTIG"
repeat_region	complement(9069..9144)
repeat_region	/rpt_family="WLTIG"
repeat_region	complement(9254..9351)
repeat_region	/rpt_family="WIR"
repeat_region	complement(9537..9688)
repeat_region	/rpt_family="MERSA"
repeat_region	9797..9942
repeat_region	/rpt_family="WIR"
repeat_region	complement(10403..10429)
repeat_region	/rpt_family="AT-rich"
repeat_region	complement(10673..10705)
repeat_region	/rpt_family="(CA)n"
repeat_region	complement(11460..11755)
repeat_region	/rpt_family="AluSx"

repeat_region	11977..12125
repeat_region	/rpt_family="MERSA"
repeat_region	12556..13643
repeat_region	/rpt_family="LIMD2"
repeat_region	13618..13808
repeat_region	/rpt_family="LIMD3"
repeat_region	complement(13997..14297)
repeat_region	/rpt_family="MER4D"
repeat_region	14327..14625
repeat_region	/rpt_family="AluSg"
repeat_region	complement(14648..14915)
repeat_region	/rpt_family="MER4D"
repeat_region	complement(14915..15147)
repeat_region	/rpt_family="MER4D"
repeat_region	15286..15404
repeat_region	/rpt_family="FLAM_C"
repeat_region	16169..16463
repeat_region	/rpt_family="AluSx"
repeat_region	16520..16570
repeat_region	/rpt_family="L2"
repeat_region	16888..16904
repeat_region	/rpt_family="L2"
repeat_region	complement(16955..17147)
repeat_region	/rpt_family="MIR"
repeat_region	complement(17171..17467)
repeat_region	/rpt_family="Aluy"
repeat_region	complement(20101..20241)
repeat_region	/rpt_family="MIR"
STS	22625..22788
	/standard_name="DXS1097"
repeat_region	/db_xref="GDB:190032"
repeat_region	24144..24435
repeat_region	/rpt_family="AluSq"
repeat_region	complement(25010..25235)
repeat_region	/rpt_family="L2"
repeat_region	25643..25771
repeat_region	/rpt_family="MERS8A"
repeat_region	complement(26111..26429)
repeat_region	/rpt_family="L2"
repeat_region	complement(26578..26636)
repeat_region	/rpt_family="(CAT)n"
repeat_region	complement(26956..27027)
repeat_region	/rpt_family="LIM3"
repeat_region	complement(27038..27188)
repeat_region	/rpt_family="L2"
repeat_region	27203..27641
repeat_region	/rpt_family="LIM4"
repeat_region	27642..28179
repeat_region	/rpt_family="LIP4"
repeat_region	28194..28455
repeat_region	/rpt_family="AluSx"
repeat_region	complement(28509..28660)
repeat_region	/rpt_family="L2"
repeat_region	28665..28830
repeat_region	/rpt_family="Aluub"
repeat_region	complement(28848..29503)
repeat_region	/rpt_family="L2"
Query Match	100.0%; Score 137; DB 9; Length 98056;
Best Local Similarity	100.0%; Pred. No. 4.6e-28;
Matches 137; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 ATTATAGGAAGAATAACCAATGCACAGTGGTGAACTGTGACTCAGGTGT 60
Db	
91659	ATTATAGGAAGAATAACCAATGCACAGTGGTGAACTGTGACTCAGGTGT 91718
QY	61 GCACAATTATCAGGAACCCCAACCAATGCAGTAGAGTAAGATAGCAGAGCGGTG 120
Db	
91719	GCACAATTATCAGGAACCCCAACCAATGCAGTAGAGTAAGATAGCAGAGCGGTG 91778
QY	121 TTTCATTTAAATAATT 137
Db	
91779	TTTCATTTAAATAATT 91795

  

RESULT 11	AR220819/c	13977 bp	DNA	linear	PAT 26-SEP-2002
LOCUS	AR220819	Sequence 60 from patent US 6426186.			
DEFINITION	AR220819	GI:23327696			
ACCESSION	AR220819	GI:23327696			
VERSION	AR220819.1	GI:23327696			
KEYWORDS	Unknown.				
SOURCE	Unknown.				
ORGANISM	Unclausified.				
REFERENCE	1 (bases 1 to 13977)				
AUTHORS	Jones,K.A., Volkmut,W. and Walker,M.G.				
TITLE	Bone remodeling genes				
JOURNAL	Patent: US 6426186-A 60 30-JUL-2002;				
FEATURES	Location/Qualifiers				
source	1..13977				
	/organism="unknown"				
	/mol_type="genomic DNA"				
ORIGIN					
Query Match	92.0%; Score 126; DB 6; Length 13977;				
Best Local Similarity	99.3%; Pred. NO. 6.9e-25;				
Matches 137; Conservative	0; Mismatches 0; Indels 1; Gaps 1;				
QY	1 ATTATAGGAAGAATAACCAATGCACAGTGGTGAACTGTGAACTCAGGTGT 60				
Db					
13306	ATTATAGGAAGAATAACCAATGCACAGTGGTGAACTGTGAACTCAGGTGT 13247				
QY	61 GCACAATTATCAGGAACCCCAACCAATGCAGTAGAGTAAGATAGCAGAG-CGGT 119				
Db					
13246	GCACAATTATCAGGAACCCCAACCAATGCAGTAGAGTAAGATAGCAGAGCCCGT 13187				
QY	120 GTTTGATGTTAATAATT 137				
Db					
13186	TTTTGATGTTAATAATT 13169				
RESULT 12	AF070485/c	13887 bp	mRNA	linear	MAM 09-DEC-1998
LOCUS	AF070485	Canis familiaris dystrophin mRNA, complete cds.			
DEFINITION	AF070485	GI:3982750			
ACCESSION	AF070485	GI:3982750			
VERSION	AF070485.1	GI:3982750			
KEYWORDS	Canis familiaris (dog)				
SOURCE	Canis familiaris				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.				
REFERENCE	1 (bases 1 to 13887)				
AUTHORS	Carville,K.S., Mann,C.J., Schatzberg,S.J. and Wilton,S.D.				
TITLE	Direct Submission				
JOURNAL	Submitted (04-JUN-1998) ANRI, Pathology, University of Western Australia, Verdun Street, Nedlands, WA 6018, Australia				
FEATURES	Location/Qualifiers				
source	1..13887				
	/organism="Canis familiaris"				
	/mol_type="mRNA"				

```

LTTETANLSDYQTALEBEVLWLLSADLALQAGGTSNDVEVKQFQTHHGYMDLTS
HGRVGNVLQSGILGTGKJSEDETEVQOMNLLNREWECLRVASMEKOSNHLKVL
MDLQNLKELNDLTKTEERKQVEPLGPDIEDLKRQVQKHVQLEQDEQVQVRL
NSLTHVWVVDSSGCHATAALEQKVLGRWANI CRWTEDRWVLLQDILLKQRET
EQCLFSAWLSKEDAVNKIHTTGFQDQSEVLNQLAVLTKDLEKKQTMKQCLSL
NODLSALKNTVAHMAWLDNPAORWDLNVLQLEKSSAOISQAVTTSPSLQTTVV
METVMTVTRHLLVKAQEEPLPPPOKQKQIIVDSIRKRLVDITLHSMITRSE
AVLOSPEFALYKKEGNSDLKKNVAIBREKAEKFKLQDASRSQAALVEQVNEGVN
ADSIKQASQLNSRWIEFCOLLSENLNLEVOYNNITFYNQQLQEQEPMFNLWLT
QPTTTSFPAISQKICKENIRLSALOPOERLKIQSIALKKQGPMLDADFVA
FTNHNQVADYQAREKELQTFDLPMPVQETSTLTITWQSETKLSLPQVTVTE
YDIMEQRLQELQSGVQEQQLNYSITVQKMSKAPLSDISRKQSFEEIEGR
WKLQSLQVHCKQLEBOWAKRLQNHKIKTIVTDVFLKEENPALGDSBILKR
OLKQCELLNDIOTQPSLNSVNEGAQKMEAPPEAFAGLETELRELNTDQWYMCQ
VYAREALKGLDKTVSLOKDLSEMHENWTOAEYLERDEFEYKTPDQLQAVEMKR
AKEAQKQKAKLSTVNSVIAQAPFAAKKALKEIDITFNMICITLKNCKCT
LEEVWACHHLLSYLEKANKWLSVEVEVKLTNTENSGABEIAEVLDSLEMLHSHED
NPNQRIITAQITIDGQWDELINLEETFNSRRELHEEAVRRKLLSEOSKQAEIE
KSLHITQESLISIDQLAAYIADKYAQMPOEAQKIOSDLTSHESISELSEKQNGE
VQEVVOSLOHCVNLVYKLSSEVKSEVMVKTGRQIVQKOTENPKELDERVPALETKS
ETAQVLSIOIDVAAKQVDSMKFRLFOKPFANFEORLOESKMILDEVKMLPALETKS
VQEVVOSLOHCVNLVYKLSSEVKSEVMVKTGRQIVQKOTENPKELDERVPALETKS
HYNELGAKTTERKQLEKCLKLSRKREKEMALTEWLAATDMELTKRAVGMENSLD
SEVANGAKTQKIEQKQVHLKSVTEVGAALKTVIGKKEMVLEDKLSLNSLNWIAVTSR
ABEWNLLEYLQKMETFPQNDVYITNMIIOADLDESEKXKPOCKEDILKRAISPLKE
NDIRPKVSTRQAAANLMAHGDHCKVVEPKISLHNRFAAISHRITKQASIPLKE
LEQFNSDIQKLEPLEAEITQOQVNLKEEDFNKMSDENEGTVKELLQKQDNLQRIED
BRKREIKTKQOLLQTKHNAKLSORRKALETSHOWYQVROADLLKLDQDIEK
KLASUPEPDERKIKELDELOKKEELNAVERQAEGLSEGAQAVAPETOIOLSKRW
RIEBSKPAQPRINFAQIHTVEESVAVKTEMPLEISYVSTVTEITHVSQALSEV
SELLNAPDCAQDFEDFKQESLKNKDSLOQISGRIDITHNKTAALHATPAERA
KLOEALSRDLDFQWERNVNRKQGRFORSVEKRRFYDMKILNQMTAEQFLKKT
QIPENWHAQYKWLKELQDGGIQSQSVVRLNATGERIIQOOSKTDASILQELKLSL
NLRWQVCHQALAEKRLQEQNLSIEFQDVNEFVLMLEADNVANIPLPQNEQQL
KEKLEOVKLAELPLQOILKOLNETGTVLVSAPLSPLEODKLENKLOKNTQWIK
VSRNLEPEKOEIEEAVKQDLOGLEOLNHLILLWSPIRNOLEIYNOPNOTGPDLEKEI
VAVQAKOPVEGILSGDGLYKXEPATOPAKEKLEDLSLDMKVVTLLOELRAKQOPG
AGLTITVRAPQGTVLVTPQAVTKETALSKLEMFSSLLLEVPALAFNRWLTLDW
LSLLDVRKTSQWMDLEDEINEMIIKOKATLODLEQRPOLEELITAAQNLKNTSN
QEARTIITPRIEIOSQDEVOHQNLQRLQTEMKDSOTWLEAKEBAEQVLOQARA
KLESWEKAPYVDIAQKITETKQAKOLRQWQVINDVANDLALXLDYADDTQV
HMTENINASWASIKHRSEREALEBETHRLLOQFPLDLEKFLAWLTAETVLANVQ
ATHKBLRDSQKQWODLOGEITEAHTDIYHNDENGQKVLRSLEGSDDAALL
SPNKVPYIYNHETQTCWDPKMTLYQSLADLNVRFSAYRTAMKRLRLOKALCLDL
LSLSACADALDOHNLKONDQMDILOVINCLTIIYDRLEQEHNLVNVPLQVDMCLNW
LLNVDTGTGRIVLSFKTGIISLCKAHLEDKRYLFOKVASSTGFCDDORLGLLH
DSIQTPROQGEVASGGSNLEPVSRCQFANKEPELSAQLDWMRLPQSMVWLPV
LHRVAAETAKQAKNICKCPIIGFRISLKNFNIDICQSCFFSGRVAQKHMHP
MVEYCTPTTSGEDVRDFAVLKKNKPKTKYFAKHPMNGVLPVTVLEGDNMETPVTLI
NFWPVDSPASSPQSLSHDTHRIEYASRLKMNMSGVLYLDSINENSDDEHLL
IQHYWRLSNQSPSPQSPQALILISLESEERGELEILADLEGRNRLQAEYDLKQ
QHEHGLSLPQSPGPMPTSPQSDAEILAEAKLRQKRLRLEARMOLLEIDHKNQLE
SOLHRLQOLLEQPARAKVNGTIVSSPSTLSORDSOPMLLRVVGVSQTSSESMEEDL
LAPQDQTSGLSEVMEQLNHSFPSSGRNVTPOKPNREDTM"
3'UTR                                     11333..11387
repeat_region                             11789..11828
/rpt_type=tandem
/rpt_unit="Ca"

ORIGIN
Query Match          75.0%; Score 102.8; DB 4; Length 13887;
Best Local Similarity 86.9%; Pred. No. 2e-18;
Matches 113; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 4 ATAAAGAAAAAAGAAATACGCAATGGAAGTGTGTAAGTGTGTAATCAGGTGTGCA 63
Db 13264 AAAATGAGAAAAATGTTACATGATGAGACAGTAGTAGGCTGTGAAATCAGGTGTGCA 13205
QY 64 CAATTATCAGGAACACCCCAAAACCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGTGTT 123

Db 13204 CGATTATAGGAACACCCCAAAACCAAGTGAAGTGAAGTGAAGTGAAGTGTGTT 13145
QY 124 GATGTTAAAT 133
Db 13144 GATGTTAAAT 13135

RESULT 13
AX306153/c
LOCUS AX306153 13815 bp DNA linear PAT 11-DEC-2001
DEFINITION Sequence 904 from Patent WO0188188.
ACCESSION AX306153
VERSION AX306153.1 GI:117645441
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Ishikawa, K., Asai, S., Takahashi, Y., Nagata, T. and Ishii, Y.
TITLE Method for examining ischemic conditions
JOURNAL Patent: WO 0188188-A 904 22-NOV-2001;
SCHOOL Juridical Person Nihon University (JP)
FEATURES
source
1. 13815
/organism="Mus musculus"
/mol_type="unassigned DNA"
/db_xref="taxon:10090"
ORIGIN
Query Match          73.4%; Score 100.6; DB 6; Length 13815;
Best Local Similarity 85.5%; Pred. No. 8.2e-18;
Matches 112; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 4 ATAAAGAAAAAAGAAATACGCAATGGAAGTGTGGAAGTGTGAACTCAGGTGTGCA 63
Db 13193 AAAAGGAAACTGTGTCCACACACAGCAAGTGTGAAAGTGTGAATAGGTGTGCA 13134
QY 64 CAATTATCAGGAACACCCCAAAACCAAGTGAAGTGAAGTGAAGTGAAGTGTGTT 123
Db 13133 CAATTACTAGAAACACCCCAAAACCAAGTGAAGTGAAGTGAAGTGTGTT 13074
QY 124 GATGTTAAATTA 134
Db 13073 GATGTTAAATTA 13063

RESULT 14
AX338582/c
LOCUS AX338582 13815 bp DNA linear PAT 23-NOV-2002
DEFINITION Sequence 2 from Patent WO0229056.
ACCESSION AX338582
VERSION AX338582.1 GI:25271088
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Chamberlain, J.S. and Harper, S.Q.
TITLE Mini-dystrophin nucleic acid and peptide sequences
JOURNAL Patent: WO 0229056-A 2 11-APR-2002;
THE REGENTS OF THE UNIVERSITY OF MICHIGAN (US)
FEATURES
source
1. 13815
/organism="Mus musculus"
/mol_type="unassigned DNA"
/db_xref="taxon:10090"
ORIGIN
Query Match          73.4%; Score 100.6; DB 6; Length 13815;
Best Local Similarity 85.5%; Pred. No. 8.2e-18;
Matches 112; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 4 ATAAAGAAAAAAGAAATACGCAATGGAAGTGTGGAAGTGTGAACTCAGGTGTGCA 63
Db 13264 AAAATGAGAAAAATGTTACATGATGAGACAGTAGTAGGCTGTGAAATCAGGTGTGCA 13205
QY 64 CAATTATCAGGAACACCCCAAAACCAAGTGAAGTGAAGTGAAGTGAAGTGTGTT 123
```





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 11:33:41 ; Search time 342.077 Seconds  
(without alignments)  
1500.955 Million cell updates/sec

Title: US-09-966-264D-1

Perfect score: 137  
Sequence: 1 attataaggaagaaagaaaa.....gtgttgatgtaattaatt 137

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2470632 seqs, 1873875610 residues

Total number of hits satisfying chosen parameters: 4941264

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_NA.\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US05\_PUBCOMB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US04\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US03\_PUBCOMB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/US02\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US01\_PUBCOMB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US00\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US05\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US04\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/1/pubpna/US03\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/1/pubpna/US02\_PUBCOMB.seq.\*
- 17: /cgn2\_6/ptodata/1/pubpna/US01\_PUBCOMB.seq.\*
- 18: /cgn2\_6/ptodata/1/pubpna/US00\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	137	100.0	137	9	US-09-966-264-1 Sequence 1, Appli
2	137	100.0	430	9	US-09-796-692-3505 Sequence 3505, Ap
3	137	100.0	430	14	US-10-040-862-3505 Sequence 3505, Ap
4	137	100.0	430	15	US-10-057-475B-3505 Sequence 3505, Ap
5	137	100.0	430	15	US-10-154-884B-3505 Sequence 3505, Ap
6	137	100.0	986	9	US-09-966-264-2 Sequence 2, Appli
7	137	100.0	2691	15	US-10-149-736-38 Sequence 38, Appli
8	137	100.0	8689	15	US-10-149-736-42 Sequence 42, Appli
9	137	100.0	11443	15	US-10-149-736-44 Sequence 44, Appli
10	137	100.0	13957	9	US-09-782-378A-22 Sequence 22, Appli
11	137	100.0	13957	9	US-09-880-107-2284 Sequence 2284, Ap
12	137	100.0	13957	15	US-10-149-736-1 Sequence 1, Appli
13	137	100.0	14089	12	US-10-342-887-434 Sequence 434, App
14	137	100.0	14082	12	US-10-342-887-981 Sequence 981, App
15	137	100.0	14082	15	US-10-341-434-108 Sequence 108, App

C 16	100.6	73.4	13815	15	US-10-149-736-2	Sequence 2, Appli
C 17	31.6	23.1	292	12	US-10-424-593-138612	Sequence 138612,
C 18	31.6	23.1	663	15	US-10-027-632-199793	Sequence 199793,
C 19	31.6	23.1	2000	9	US-09-938-842A-5066	Sequence 5066, Ap
C 20	31.6	23.1	2000	11	US-09-938-842A-5066	Sequence 5066, Ap
C 21	31.2	22.8	1977	10	US-09-873-367C-994	Sequence 994, App
C 22	31.2	22.8	1977	12	US-10-240-425-1268	Sequence 1268, Ap
C 23	31.2	22.8	1977	12	US-10-342-887-382	Sequence 382, App
C 24	31.2	22.8	2463	14	US-10-240-965-100	Sequence 100, App
C 25	30.8	22.5	1157	15	US-10-027-632-208705	Sequence 208705,
C 26	30.8	22.5	36303	14	US-10-152-724A-24	Sequence 24, Appli
C 27	30.6	22.3	538	9	US-09-867-701-1432	Sequence 1432, App
C 28	30.6	22.3	539	9	US-09-917-800A-153	Sequence 153, App
C 29	30.4	22.2	613	15	US-10-027-632-105562	Sequence 105562,
C 30	30.4	22.2	613	15	US-10-027-632-126516	Sequence 126516,
C 31	30.4	22.2	624	15	US-10-027-632-106141	Sequence 106141,
C 32	30.4	22.2	704	15	US-10-027-632-17066	Sequence 17066, A
C 33	30.4	22.2	763	15	US-10-027-632-126517	Sequence 126517,
C 34	30.4	22.2	786	15	US-10-027-632-126515	Sequence 126515,
C 35	30	21.9	4974	9	US-09-906-209-17	Sequence 17, Appli
C 36	30	21.9	6611	10	US-09-934-900-26	Sequence 26, Appli
C 37	30	21.9	7701	15	US-10-459-159-1	Sequence 1, Appli
C 38	30	21.9	7701	15	US-10-427-570A-9	Sequence 9, Appli
C 39	30	21.9	99014	9	US-09-880-107-3428	Sequence 3428, Ap
C 40	29.8	21.8	1015	14	US-10-259-165-589	Sequence 589, App
C 41	29.8	21.8	67191	14	US-10-105-612-1	Sequence 1, Appli
C 42	29.6	21.6	277	12	US-10-424-599-44445	Sequence 44445, A
C 43	29.6	21.6	382	12	US-10-424-599-45074	Sequence 45074, A
C 44	29.4	21.5	575	9	US-09-920-300A-4	Sequence 4, Appli
C 45	29.4	21.5	575	13	US-10-033-528-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1  
US-09-966-264-1  
; Sequence 1, Application US/09966264  
; Patent No. US20020099015A1  
; GENERAL INFORMATION:  
; APPLICANT: Barber, Elizabeth K  
; TITLE OF INVENTION: Gene Expression Control Element DNA  
; FILE REFERENCE: 896034605001  
; CURRENT APPLICATION NUMBER: US/09/966,264  
; CURRENT FILING DATE: 2001-09-28  
; PRIOR APPLICATION NUMBER: US 60/237,079  
; PRIOR FILING DATE: 2000-09-30  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 137  
; TYPE: DNA  
; ORGANISM: human  
; FEATURE:  
; NAME/KEY: exon  
; LOCATION: (3)..(137)  
; OTHER INFORMATION:  
; NAME/KEY: PolyA\_site  
; LOCATION: (130)..(135)  
; OTHER INFORMATION:  
US-09-966-264-1

Query Match	100.0%	Score 137;	DB 9;	Length 137;
Best Local Similarity	100.0%;	Pred. No. 2e-33;		
Matches 137;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ATTATAAGGAAAAAGAAATAACGCAATGGAACGCTGGAAGTCACTGAGTGT	60	
Db	1	ATTATAAGGAAAAAGAAATAACGCAATGGAACGCTGGAAGTCACTGAGTGT	60	
QY	61	GCACATTTATCAGGAACACCCCAAAACCAAGTCAGGTAGTAATAGCATGAGACCGGTG	120	
Db	61	GCACATTTATCAGGAACACCCCAAAACCAAGTCAGGTAGTAATAGCATGAGACCGGTG	120	





```

; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 11290
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3505
; LENGTH: 430
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-154-884B-3505

Query Match          100.0%; Score 137; DB 15; Length 430;
Best Local Similarity 100.0%; Pred. No. 3.1e-33;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATTAAGGMAAAAGAAATAACGCAATGCGACAGTGTGAAGCTGTGAACCTCAGGTGT 60
Db 287 ATTATTAAGGMAAAAGAAATAACGCAATGCGACAGTGTGAAGCTGTGAACCTCAGGTGT 228
QY 61 GCACAAATTATCAGGACACACCCCAAAACCAAGTAGAGGTAGCAATAGCATGAGAAGCCGTG 120
Db 227 GCACAAATTATCAGGACACACCCCAAAACCAAGTAGAGGTAGCAATAGCATGAGAAGCCGTG 168
QY 121 TTGTGATGTTAAATAATT 137
Db 167 TTGTGATGTTAAATAATT 151

RESULT 6
US-09-366-264-2
; Sequence 2, Application US/09966264
; Patent No. US2002009015A1
; GENERAL INFORMATION:
; APPLICANT: Barber, Elizabeth K
; TITLE OF INVENTION: Gene Expression Control Element DNA
; FILE REFERENCE: 896034605001
; CURRENT APPLICATION NUMBER: US/09/966,264
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US 60/237,079
; PRIOR FILING DATE: 2000-09-30
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 996
; TYPE: DNA
; ORGANISM: human
; FEATURE:
; NAME/KEY: exon
; LOCATION: (1)...(996)
; OTHER INFORMATION:
; NAME/KEY: misc_feature
; LOCATION: (710)...(996)
; OTHER INFORMATION: Nucleotides 710-996 are homologous to a portion of human
; OTHER INFORMATION: in DNA in the region of exon 79 except that nucleotides
; OTHER INFORMATION: re inverted in comparison to the orientation of the same
; OTHER INFORMATION: in the dystrophin DNA
US-09-366-264-2

Query Match          100.0%; Score 137; DB 9; Length 996;
Best Local Similarity 100.0%; Pred. No. 4.3e-33;

```

	Matches	137;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1	ATTATAAGGAAAAAGAAAAAT	AAGC	CAATG	GCACAA	GTGTGA	AGCTGTG	GA	ACTCAGG	TGT 60
Db	860	ATTATAAGGAAAAAGAAAAAT	AAGC	CAATG	GCACAA	GTGTGA	AGCTGTG	GA	ACTCAGG	TGT 919
QY	61	GCACAA	TATT	CAGGA	ACACCC	CAAAAC	CAAAGT	GAGT	AGAAAT	AGCATG
Db	920	GCACAA	TATT	CAGGA	ACACCC	CAAAAC	CAAAGT	GAGT	AGAAAT	AGCATG
QY	121	TTTGAT	GTGTTAA	TTAATT	137					
Db	980	TTTGAT	GTGTTAA	TTAATT	996					

```

RESULT 7
US-10-149-736-38/c
: Sequence 38, Application US/10149736
: Publication No. US20030216332A1
: GENERAL INFORMATION:
: APPLICANT: Chamberlain, Jeffrey S.
: APPLICANT: Harper, Scott Q.
: TYPE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences

```

```

? FILE REFERENCE: UM-U0968
? CURRENT APPLICATION NUMBER: US/10/149,736
? CURRENT FILING DATE: 2002-06-17
? PRIOR APPLICATION NUMBER: PCT/US01/31126
? PRIOR FILING DATE: 2001-10-04
? PRIOR APPLICATION NUMBER: 60/238,848
? PRIOR FILING DATE: 2000-10-06
? NUMBER OF SEQ ID NOS: 96
? SOFTWARE: PatentIn version 3.1
? SEQ ID NO 38
? LENGTH: 2891
? TYPE: DNA
? ORGANISM: Homo sapiens
? US-10-149-736-38

```

	Query Match	100.0%;	Score 137;	DB 15;	Length 2691;
	Best Local Similarity	100.0%;	Pred. No. 6.3e-33;		
	Matches 137;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	ATTATAAGGAAAAAGAAAATAACGCAATGGACAAAGTGGTGAAGCTGTGAACTCGAGGT	60		
Db	2033	ATTATAAGGAAAAAGAAAATAACGCAATGGACAAAGTGGTGAAGCTGTGAACTCGAGGT	1974		
Qy	61	GCACAAATTATCAGGAACACCCCAAAACCCAAAGTCAGGTAGAAATAGCATGAGAAGCCCGT	120		
Db	1973	GCACAAATTATCAGGAACACCCCAAAACCCAAAGTCAGGTAGAAATAGCATGAGAAGCCCGT	1914		
Qy	121	TTTGATGTTAAATTAAAT	137		
Db	1913	TTTGATGTTAAATTAAAT	1897		

```

RESULT 8
US-10-149-736-42/c
; Sequence 42, Application US/10149736
; Publication No. US20030216332A1
; GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; APPLICANT: Harper, Scott Q.
; TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences
; FILE REFERENCE: UM-06968
; CURRENT APPLICATION NUMBER: US/10/149,736
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: PCT/US01/31126
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,848
; PRIOR FILING DATE: 2000-10-06
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 42

```

```

; LENGTH: 8689
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
; US-10-149-736-42

Query Match          100.0%;   Score 137;   DB 15;   Length 8689;
Best Local Similarity 100.0%;   Pred. No. 9.8e-33;
Matches 137;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy      1  ATTATTAAGGAAAAAGAAAATAACCCATGACCAAGTGGTGAAGGTGTGAACTCAGGTGT 60
Db      8031 ATTATTAAGGAAAAAGAAAATAACCCATGACCAAGTGGTGAAGTGTGAACTCAGGTGT 7972

Qy      61  GCACAAATTATCAGGAACACCCCAAAACCAAAAGTGAAGTATGAGTAATGAGTGAAGCCGCTG 120
Db      7971 GCACAAATTATCAGGAACACCCCAAAACCAAAAGTGAAGTATGAGTAATGAGTGAAGCCGCTG 7912

Qy      121  TTTTGATGTTAAATAATT 137
Db      7911 TTTTGATGTTAAATAATT 7895

```

RESULT 9  
US-10-149-736-44/c  
; Sequence 44, Application US/10-49736  
; Publication No. US20030216332A1  
; GENERAL INFORMATION:  
; APPLICANT: Chamberlain, Jeffrey S.  
; APPLICANT: Harper, Scott Q.  
; TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences  
; FILE REFERENCE: UM-06968  
; CURRENT APPLICATION NUMBER: US/10/149,736

```

, CURRENT FILING DATE: 2002-06-17
,
, PRIOR APPLICATION NUMBER: PCT/US01/31126
,
, PRIOR FILING DATE: 2001-10-04
,
, PRIOR APPLICATION NUMBER: 60/236,848
,
, PRIOR FILING DATE: 2000-10-06
,
, NUMBER OF SEQ ID NOS: 96
,
, SOFTWARE: PatentIn version 3.1
,
, SEQ ID NO 44
,
, LENGTH: 11443
,
, TYPE: DNA
,
, ORGANISM: Artificial Sequence
,
, FEATURE:
,
, OTHER INFORMATION: Synthetic
,
, US-10-149-736-44

```

	Query Match	100.0%;	Score 137;	DB 15;	Length 11443;
	Best Local Similarity	100.0%;	Pred. No. 1.1e-32;		
	Matches 137;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ATTATTAAGGAAAAGAAATTAACGCAATGGGACAAAGTGGTGAAGCTGTGAACTCAGGTTG	60		
DB	10785	ATTATTAAGGAAAAGAAATTAACGCAATGGGACAAAGTGGTGAAGCTGTGAACTCAGGTTG	10726		
QY	61	GCACAAATTATCAGGACACACCCCAAAACCAAGGTGAGTAATAGCATGAGAAGCCGTTG	120		
DB	10725	GCACAAATTATCAGGACACACCCCAAAACCAAGGTGAGTGAATAGCATGAGAAGCCGTTG	10656		
QY	121	TTTGAAGTTAATTAAATT	137		
DB	10665	TTTGATGTTTAATTAAATT	10649		

RESULT 10  
US-09-782-378A-22/c  
; Sequence 22, Application US/09782378A  
; Patent No. US20020102731A1  
; GENERAL INFORMATION:  
; APPLICANT: Hearing, Patrick  
; APPLICANT: Bahou, Wadie

```

; APPLICANT: Sandalon, Ziv
; APPLICANT: Gnatenko, Dmitri
; TITLE OF INVENTION: Adenoviral Vectors
; FILE REFERENCE: STONYB-04970
; CURRENT APPLICATION NUMBER: US/09/782,378A
; PRIOR FILING DATE: 2001-02-12
; PRIOR APPLICATION NUMBER: 60/237,747
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 22
; LENGTH: 13957
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-782-378A-22

Query Match 100.0%; Score 137; DB 9; Length 13957;
Best Local Similarity 100.0%; Pred. No. 1.2e-32;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATAAGGAAAAAGAAATAACGCAATGGACAAGTGGTGAAGCTGTGAAGCTCAAGTGT 60
DB 13299 ATTATAAGGAAAAAGAAATAACGCAATGGACAAGTGGTGAAGCTGTGAAGCTCAAGTGT 13240
QY 61 GCACATTATCAGGAACACCCCAACCAAGTGAAGTAGAATAAGCATGAGAGCCCGT 120
DB 13239 GCACATTATCAGGAACACCCCAACCAAGTGAAGTAGAATAAGCATGAGAGCCCGT 13180
QY 121 TTGTGTTTAATTAATT 137
DB 13179 TTGTGTTTAATTAATT 13163

RESULT 11
US-09-880-107-2284/c
; Sequence 2284, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Parci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2284
; LENGTH: 13957
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 M18533
US-09-880-107-2284

Query Match 100.0%; Score 137; DB 9; Length 13957;
Best Local Similarity 100.0%; Pred. No. 1.2e-32;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATAAGGAAAAAGAAATAACGCAATGGACAAGTGGTGAAGCTGTGAAGCTCAAGTGT 60
DB 13299 ATTATAAGGAAAAAGAAATAACGCAATGGACAAGTGGTGAAGCTGTGAAGCTCAAGTGT 13240
QY 61 GCACATTATCAGGAACACCCCAACCAAGTGAAGTAGAATAAGCATGAGAGCCCGT 120
DB 13239 GCACATTATCAGGAACACCCCAACCAAGTGAAGTAGAATAAGCATGAGAGCCCGT 13180
QY 121 TTGTGTTTAATTAATT 137
DB 13179 TTGTGTTTAATTAATT 13163

RESULT 12
US-10-149-736-1/c
; Sequence 1, Application US/10149736
; Publication No. US20030216332A1
; GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; APPLICANT: Harper, Scott Q.
; TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences
; FILE REFERENCE: UM-06968
; CURRENT APPLICATION NUMBER: US/10/149,736
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: PCT/US01/31126
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,848
; PRIOR FILING DATE: 2000-10-06
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 1
; LENGTH: 13957
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-149-736-1

Query Match 100.0%; Score 137; DB 15; Length 13957;
Best Local Similarity 100.0%; Pred. No. 1.2e-32;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATAAGGAAAAAGAAATAACGCAATGGACAAGTGGTGAAGCTGTGAAGCTCAAGTGT 60
DB 13299 ATTATAAGGAAAAAGAAATAACGCAATGGACAAGTGGTGAAGCTGTGAAGCTCAAGTGT 13240
QY 61 GCACATTATCAGGAACACCCCAACCAAGTGAAGTAGAATAAGCATGAGAGCCCGT 120
DB 13239 GCACATTATCAGGAACACCCCAACCAAGTGAAGTAGAATAAGCATGAGAGCCCGT 13180
QY 121 TTGTGTTTAATTAATT 137
DB 13179 TTGTGTTTAATTAATT 13163

RESULT 13
US-10-342-887-434/c
; Sequence 434, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/299,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 434
; LENGTH: 14069
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-434
```

Query Match 100.0%; Score 137; DB 12; Length 14069;  
Best Local Similarity 100.0%; Pred. No. 1.2e-32;  
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATTATAAGGAAAAAGAAATAACGCAATGGCAAGTGGTGAAGTGTGAATCTCAGGTGT 60  
DB 13411 ATTATAAGGAAAAAGAAATAACGCAATGGCAAGTGGTGAAGTGTGAATCTCAGGTGT 13352  
QY 61 GCACAAATTATCAGGAACACCCCAACCAACCAAGTGGTGAAGTGTGAATCTCAGGTGT 120  
DB 13351 GCACAAATTATCAGGAACACCCCAACCAACCAAGTGGTGAAGTGTGAATCTCAGGTGT 13292  
QY 121 TTTCATGTTTAATTAATT 137  
DB 13291 TTTCATGTTTAATTAATT 13275

RESULT 14  
US-10-342-887-981/c  
; Sequence 981, Application US/10342887  
; Publication No. US20040058340A1  
; GENERAL INFORMATION:  
; APPLICANT: Dai, Hongyue  
; APPLICANT: He, Yudong  
; APPLICANT: Linsley, Peter S.  
; APPLICANT: Mao, Mao  
; APPLICANT: Roberts, Christopher J.  
; APPLICANT: Van 't Veer, Laura Johanna  
; APPLICANT: Van de Vijver, Marc J.  
; APPLICANT: Bernards, Rene  
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients  
; FILE REFERENCE: 9301-188-999  
; CURRENT APPLICATION NUMBER: US/10/342,887  
; CURRENT FILING DATE: 2003-01-15  
; PRIOR APPLICATION NUMBER: 60/298,918  
; PRIOR FILING DATE: 2001-06-18  
; PRIOR APPLICATION NUMBER: 60/380,710  
; PRIOR FILING DATE: 2002-05-14  
; PRIOR APPLICATION NUMBER: 10/172,118  
; PRIOR FILING DATE: 2002-06-14  
; NUMBER OF SEQ ID NOS: 2699  
; SEQ ID NO 981  
; LENGTH: 14082  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-342-887-981

Query Match 100.0%; Score 137; DB 12; Length 14082;  
Best Local Similarity 100.0%; Pred. No. 1.2e-32;  
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATTATAAGGAAAAAGAAATAACGCAATGGCAAGTGGTGAAGTGTGAATCTCAGGTGT 60  
DB 13424 ATTATAAGGAAAAAGAAATAACGCAATGGCAAGTGGTGAAGTGTGAATCTCAGGTGT 13365  
QY 61 GCACAAATTATCAGGAACACCCCAACCAACCAAGTGGTGAAGTGTGAATCTCAGGTGT 120  
DB 13364 GCACAAATTATCAGGAACACCCCAACCAACCAAGTGGTGAAGTGTGAATCTCAGGTGT 13305  
QY 121 TTTCATGTTTAATTAATT 137  
DB 13304 TTTCATGTTTAATTAATT 13288

RESULT 15  
US-10-341-434-108/c  
; Sequence 108, Application US/10341434  
; Publication No. US20030215835A1  
; GENERAL INFORMATION:  
; APPLICANT: Origene Technologies  
; TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes  
; FILE REFERENCE: 9U 204 205 R1

; CURRENT APPLICATION NUMBER: US/10/341,434  
; CURRENT FILING DATE: 2003-07-18  
; PRIOR APPLICATION NUMBER: US 60/348,164  
; PRIOR FILING DATE: 2002-01-15  
; PRIOR APPLICATION NUMBER: US 60/348,119  
; PRIOR FILING DATE: 2002-01-15  
; NUMBER OF SEQ ID NOS: 238  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 108  
; LENGTH: 14082  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (703)..(11388)  
; OTHER INFORMATION:  
US-10-341-434-108

Query Match 100.0%; Score 137; DB 15; Length 14082;  
Best Local Similarity 100.0%; Pred. No. 1.2e-32;  
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATTATAAGGAAAAAGAAATAACGCAATGGCAAGTGGTGAAGTGTGAATCTCAGGTGT 60  
DB 13424 ATTATAAGGAAAAAGAAATAACGCAATGGCAAGTGGTGAAGTGTGAATCTCAGGTGT 13365  
QY 61 GCACAAATTATCAGGAACACCCCAACCAACCAAGTGGTGAAGTGTGAATCTCAGGTGT 120  
DB 13364 GCACAAATTATCAGGAACACCCCAACCAACCAAGTGGTGAAGTGTGAATCTCAGGTGT 13305  
QY 121 TTTCATGTTTAATTAATT 137  
DB 13304 TTTCATGTTTAATTAATT 13288

Search completed: April 6, 2004, 15:01:39  
Job time : 343.077 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 08:36:01 ; Search time 15.2357 Seconds  
(without alignments)  
4990.154 Million cell updates/sec

Title: US-09-966-264D-1

Perfect score: 137

Sequence: 1 attataaggaagaaagaaa.....gtgttgatgtaattaatt 137

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA:\*

1: /cgn2\_6/ptodata/2/ina/5A COMB.seq:\*

2: /cgn2\_6/ptodata/2/ina/5B COMB.seq:\*

3: /cgn2\_6/ptodata/2/ina/6A COMB.seq:\*

4: /cgn2\_6/ptodata/2/ina/6B COMB.seq:\*

5: /cgn2\_6/ptodata/2/ina/6C COMB.seq:\*

6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	126	92.0	13977	4	US-09-484-970B-60
C 2	100.6	73.4	19307	3	US-08-836-022A-10
C 3	100.6	73.4	19307	3	US-09-427-048A-10
C 4	28.2	20.6	5326	3	US-09-338-907-124
C 5	28.2	20.6	5326	4	US-09-218-207-124
C 6	28	20.4	38844	4	US-09-734-675-3
C 7	28	20.4	84495	4	US-09-797-906-3
C 8	27.8	20.3	1266	4	US-09-543-681A-2904
C 9	27.8	20.3	2274	3	US-08-492-459-13
C 10	27.8	20.3	2274	3	US-08-423-753-13
C 11	27.8	20.3	2274	3	US-08-716-873-27
C 12	27.8	20.3	2274	3	US-09-368-431-27
C 13	27.8	20.3	2274	4	US-09-414-006-13
C 14	27.8	20.3	2274	4	US-09-447-223-13
C 15	27.8	20.3	2799	2	US-08-874-186-42
C 16	27.8	20.3	7218	1	US-08-232-463-14
C 17	27.6	20.1	1173	4	US-09-107-532A-2915
C 18	27.6	20.1	3705	5	PCT-US96-03940-7
C 19	27.6	20.1	3786	4	US-08-961-527-182
C 20	27.6	20.1	5648	5	PCT-US96-03940-1
C 21	27.6	20.1	18355	4	US-08-956-171E-67
C 22	27.4	20.0	2421	3	US-08-851-843A-51
C 23	27.4	20.0	2421	3	US-08-974-549A-218
C 24	27.4	20.0	2421	3	US-08-854-050-51
C 25	27.4	20.0	2421	4	US-09-430-323-51
C 26	27.4	20.0	2421	4	US-09-402-181B-218
C 27	27.4	20.0	2421	4	US-09-721-456-218

C 28	27.4	20.0	3459	4	US-09-016-434-1363	Sequence 1363, Ap
C 29	27.4	20.0	58407	4	US-08-916-421B-2	Sequence 2, Appli
C 30	27.2	19.9	770	4	US-09-205-258-161	Sequence 161, App
C 31	27.2	19.9	1296	4	US-09-134-001C-1420	Sequence 1420, Ap
C 32	27.2	19.9	2508	2	US-08-850-993-1	Sequence 1, Appli
C 33	27.2	19.9	13417	2	US-08-637-759B-37	Sequence 37, Appl
C 34	27.2	19.9	13417	3	US-08-871-355A-37	Sequence 37, Appl
C 35	27.2	19.9	13417	4	US-09-201-945-37	Sequence 37, Appl
C 36	26.8	19.6	888	4	US-09-134-001C-1629	Sequence 1629, Ap
C 37	26.8	19.6	1221	4	US-09-134-001C-733	Sequence 733, App
C 38	26.8	19.6	5873	1	US-07-928-464-4	Sequence 4, Appli
C 39	26.8	19.6	5873	5	PCT-US93-07347-4	Sequence 4, Appli
C 40	26.8	19.6	5890	1	US-07-928-464-3	Sequence 3, Appli
C 41	26.8	19.6	5890	1	US-07-928-464-5	Sequence 5, Appli
C 42	26.8	19.6	5890	1	US-07-928-464-6	Sequence 6, Appli
C 43	26.8	19.6	5890	5	PCT-US93-07347-3	Sequence 3, Appli
C 44	26.8	19.6	5890	5	PCT-US93-07347-5	Sequence 5, Appli
C 45	26.8	19.6	5890	5	PCT-US93-07347-6	Sequence 6, Appli

#### ALIGNMENTS

RESULT 1

US-09-484-970B-60/c

; Sequence 60, Application US/09484970B

; Patent No. 6426186

; GENERAL INFORMATION:

; APPLICANT: Jones, Karen A.

; APPLICANT: Volkmut, Wayne

; APPLICANT: Walker, Michael G.

; TITLE OF INVENTION: BONE REMODELING GENES

; FILE REFERENCE: PB-0014 US

; CURRENT APPLICATION NUMBER: US/09/484,970B

; CURRENT FILING DATE: 2000-01-18

; NUMBER OF SEQ ID NOS: 172

; SOFTWARE: PERL Program

; SEQ ID NO 60

; LENGTH: 13977

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc\_feature

; OTHER INFORMATION: Incyte ID No. 6426186 229357.11CB1

; NAME/KEY: unsure

; LOCATION: 11721-11761, 12294, 13969

; OTHER INFORMATION: a, t, c, g, or other

; US-09-484-970B-60

Query Match 92.0%; Score 126; DB 4; Length 13977;  
Best Local Similarity 99.3%; Pred. No. 4.8e-31;  
Matches 137; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy	1	ATTATTAAGGAAAGAAATACGCAATGGCAAGTGGTGAAGCTGTGAAGCTGTGAAGCTGTGAAGCTGTGT	60
Db	13306	ATTATTAAGGAAAGAAATACGCAATGGCAAGTGGTGAAGCTGTGAAGCTGTGAAGCTGTGT	13247
Qy	61	GCACATTTATCAGAACACCCCAAAACCAAAAGTAGAGTAAATAGCATGAGAAG-CCGT	119
Db	13246	GCACATTTATCAGAACACCCCAAAACCAAAAGTAGAGTAAATAGCATGAGAAGCCGT	13187
Qy	120	GTTTGATGTTTAATTAATT	137
Db	13186	GTTTGATGTTTAATTAATT	13169

RESULT 2

US-08-836-022A-10

; Sequence 10, Application US/08836022A

; Patent No. 6001557

; GENERAL INFORMATION:

; APPLICANT: Trustees of the University of Pennsylvania

; APPLICANT: Wilson, James M.

APPLICANT: Fisher, Krishna J.  
APPLICANT: Chen, Shu-Jen  
APPLICANT: Weitzman, Matthew  
TITLE OF INVENTION: Improved Adenovirus Virus and  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSES: Howson and Howson  
STREET: Spring House Corporate Cntr, P O Box 457  
CITY: Spring House  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19477  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/836,022A  
FILING DATE: 28-OCT-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/331,381  
FILING DATE: 28-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: GNVFN.008PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-540-9200  
TELEFAX: 215-540-5818  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19307 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: CDNA  
US-08-836-022A-10

Query Match 73.4%; Score 100.6; DB 3; Length 19307;  
Best Local Similarity 85.5%; Pred. No. 8e-23;  
Matches 112; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 4 ATAAAGGAAAAAATAACGCAAGTGTGAAGCTGTGAAGCTCAGGTGTGCA 63  
DB 1492 AAAAGGAAACTGTGTCCACACACGCAAGTGTGAAGTGTGAAGTGTGCA 1551

QY 64 CAATTATCAGGAACACCCCAAAACCAAGTGTGAAGTGTGAAGTGTGCA 123  
DB 1552 CAATTACTAGAAACACCCCAAAACCAAGTGTGAAGTGTGAAGTGTGTT 1611

QY 124 GATGTTAATTA 134  
DB 1612 GATGTTAATTA 1622

RESULT 3  
US-09-427-048A-10  
Sequence 10, Application US/09427048A  
Patent No. 6203375  
GENERAL INFORMATION:  
APPLICANT: Trustees of the University of Pennsylvania  
APPLICANT: Wilson, James M.  
APPLICANT: Fisher, Krishna J.  
APPLICANT: Chen, Shu-Jen  
APPLICANT: Weitzman, Matthew  
TITLE OF INVENTION: Improved Adenovirus Virus and  
METHODS OF USE THEREOF  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSES: Howson and Howson  
STREET: Spring House Corporate Cntr, P O Box 457

CITY: Spring House  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19477  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/427,048A  
FILING DATE: 21-Oct-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/836,022  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: GNVFN.008PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-540-9200  
TELEFAX: 215-540-5818  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19307 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: CDNA  
SEQUENCE DESCRIPTION: SEQ ID NO: 10:  
US-09-427-048A-10

Query Match 73.4%; Score 100.6; DB 3; Length 19307;  
Best Local Similarity 85.5%; Pred. No. 8e-23;  
Matches 112; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 4 ATAAAGGAAAAAATAACGCAAGTGTGAAGCTGTGAAGCTCAGGTGTGCA 63  
DB 1492 AAAAGGAAACTGTGTCCACACACGCAAGTGTGAAGTGTGAAGTGTGCA 1551

QY 64 CAATTATCAGGAACACCCCAAAACCAAGTGTGAAGTGTGAAGTGTGCA 123  
DB 1552 CAATTACTAGAAACACCCCAAAACCAAGTGTGAAGTGTGAAGTGTGTT 1611

QY 124 GATGTTAATTA 134  
DB 1612 GATGTTAATTA 1622

RESULT 4  
US-09-338-907-124/c  
Sequence 124, Application US/09338907  
Patent No. 6265546  
GENERAL INFORMATION:  
APPLICANT: Cohen, Daniel  
APPLICANT: Blumenfeld, Marta  
APPLICANT: Ilya, Chumakov  
APPLICANT: Bougueleret, Lydie  
TITLE OF INVENTION: PROSTATE CANCER GENE  
FILE REFERENCE: GENSET.18CP1CP  
CURRENT APPLICATION NUMBER: US/09/338,907  
CURRENT FILING DATE: 1999-06-23  
EARLIER APPLICATION NUMBER: 08/996,306  
EARLIER FILING DATE: 1997-12-22  
EARLIER APPLICATION NUMBER: 60/099,658  
EARLIER FILING DATE: 1998-09-09  
EARLIER APPLICATION NUMBER: 09/218,207  
EARLIER FILING DATE: 1998-12-22  
NUMBER OF SEQ ID NOS: 578  
SOFTWARE: Patent.pm  
SEQ ID NO 124  
LENGTH: 5324

```
;
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31..588
; NAME/KEY: polyA signal
; LOCATION: 5297..5302
; OTHER INFORMATION: AATAAA
US-09-338-907-124

Query Match      20.6%; Score 28.2; DB 3; Length 5326;
Best Local Similarity 52.8%; Pred. No. 11;
Matches 57; Conservative 1; Mismatches 50; Indels 0; Gaps 0;

QY 11 AAAAGAAAATAACCAATGACCAAGTGTGAAGCTGTGAAGTGTGAAGTGTGACAAATTAT 70
Db 661 ACAAGAGAACACGAGCTGGR.GGCGGTGGCAGCTGTATGTCTGTAGTCTTAAGTACT 602

QY 71 CAGGAACACCCCAAAACCAAGTGTAGGTTAGAAATAGCATGAGAAGCCG 118
Db 601 CAGGAGGACTGCTTTAGCCACGTTGGCGAGCAAAATGCCTGACTAGCTG 554

RESULT 5
US-09-218-207-124/c
; Sequence 124, Application US/09218207
; Patent No. 6346381
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Ilya, Chumakov
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: Prostate cancer gene
; FILE REFERENCE: GENSET.018CPI
; CURRENT APPLICATION NUMBER: US/09/218,207
; EARLIER FILING DATE: 1998-12-22
; EARLIER FILING DATE: 1997-12-22
; EARLIER FILING DATE: 1997-12-22
; EARLIER FILING DATE: 1998-09-09
; NUMBER OF SEQ ID NOS: 578
; SOFTWARE: Patent.pm
; SEQ ID NO 124
; LENGTH: 5324
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31..588
; NAME/KEY: polyA signal
; LOCATION: 5297..5302
; OTHER INFORMATION: AATAAA
US-09-218-207-124

Query Match      20.6%; Score 28.2; DB 4; Length 5326;
Best Local Similarity 52.8%; Pred. No. 11;
Matches 57; Conservative 1; Mismatches 50; Indels 0; Gaps 0;

QY 11 AAAAGAAAATAACCAATGACCAAGTGTGAAGCTGTGAAGTGTGACCAATTAT 70
Db 661 ACAAGAGAACACGAGCTGGR.GGCGGTGGCAGCTGTATGTCTGTAGTCTTAAGTACT 602

QY 71 CAGGAACACCCCAAAACCAAGTGTAGGTTAGAAATAGCATGAGAAGCCG 118
Db 601 CAGGAGGACTGCTTTAGCCACGTTGGCGAGCAAAATGCCTGACTAGCTG 554

RESULT 6
US-09-734-675-3
; Sequence 3, Application US/09734675
; Patent No. 6365391
; GENERAL INFORMATION:
; APPLICANT: WEBSTER, Marion et al

; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL000862
; CURRENT APPLICATION NUMBER: US/09/734,675
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 38844
; TYPE: DNA
; ORGANISM: Human
US-09-734-675-3

Query Match      20.4%; Score 28; DB 4; Length 38844;
Best Local Similarity 50.8%; Pred. No. 23;
Matches 67; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 1 ATTATTAAGGAAAAGAAAATAACGCAATGGCAAGTGTGAAGCTGTGAAGTGTGAACTCAGGTGT 60
Db 526 AATATGTATGAGTGAATGAATGATTAATACATTAATGATAATTTATTATTAATTTGA 585

QY 61 GCACAAATTATCAGGAACACCCCAAAACCAAGTGTAGGTTAGAAATAGCATGAGAAGCCGTG 120
Db 586 ACGAAATCAATCGAGAGGACCGAGATAAATTTGTGCTCTAGAGTAAGAAGACCTGAG 645

QY 121 TTTGATGTTAAT 132
Db 646 TTTGAGATAACT 657

RESULT 7
US-09-797-906-3/c
; Sequence 3, Application US/09797906
; Patent No. 6329188
; GENERAL INFORMATION:
; APPLICANT: Zianghe YAN, Karen A. KETCHUM, Valentina DIFRANCESCO, Ellen M. BEASLEY
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001151CIP
; CURRENT APPLICATION NUMBER: US/09/797,906
; CURRENT FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 84495
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(84495)
; OTHER INFORMATION: n = A,T,C or G
US-09-797-906-3

Query Match      20.4%; Score 28; DB 4; Length 84495;
Best Local Similarity 55.0%; Pred. No. 29;
Matches 55; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 2 TTATAAGGAAAAGAAAATAACGCAATGGCAAGTGTGAAGCTGTGAAGTGTGAACTCAGGTGTG 61
Db 9226 TTATAAGGAAAAGAAAATAACGCAATGGCAAGTGTGAAGCTGTGAAGCTGTGAAGCTGTGA 9167

QY 62 CACAATTATCAGGAACACCCCAAAACCAAGTGTAGGTTAGAAATAGCATGAGAAGCCG 101
Db 9166 CAATAAAAAAATAACGCAAGTGTGAAGCTGTGAAGCTGTGAAGCTGTGAAGCTGTGA 9127

RESULT 8
US-09-543-681A-2904/c
; Sequence 2904, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
```



```

; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 2904
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-2904

Query Match 20.3%; Score 27.8; DB 4; Length 1266;
Best Local Similarity 55.1%; Pred. No. 9;
Matches 41; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 65 AATTATCAGGACACCCCAAAAGTGGTGAATAGCATGAGAGCCGCTGTTG 124
Db 886 ATTATCAGAGAACCCAGCTAAATAAGAGTGGCCAGGATAGCAACCAAGGCATGAGTG 827

QY 125 ATG 127
Db 826 ATG 824

RESULT 9
US-08-492-459-13/c
; Sequence 13, Application US/08492459
; Patent No. 6015689
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: REGULATION OF AUROBASIDIN SENSITIVITY IN FUNGUS
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.4 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/492,459
; FILING DATE: June 20, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; FILING DATE: May 16, 1994
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2274
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
US-08-492-459-13

Query Match 20.3%; Score 27.8; DB 3; Length 2274;
Best Local Similarity 55.1%; Pred. No. 9;
Matches 41; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

```

```

Best Local Similarity 57.5%; Pred. No. 11;
Matches 50; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 50 AACTCAGGTGTGCACAATTATCAGGACACCCCAAAAGTGGTGAATAGCAT 109
Db 1848 AACACATTATGTAAAAAAGAGACACACACAAAAAAGTGGTGAATAGCAT 1789

QY 110 GAGAAGCCGCTTTGATGTTAATAAT 136
Db 1788 GCCATTCTATATTGATGTTAATAAT 1762

RESULT 10
US-08-423-752-13/c
; Sequence 13, Application US/08423752
; Patent No. 6022349
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: A GENE CODING FOR A PROTEIN REGULATING
; TITLE OF INVENTION: AUROBASIDIN SENSITIVITY
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/423,752
; FILING DATE: April 18, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/243,403
; FILING DATE: May 16, 1994
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2274
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
US-08-423-752-13

Query Match 20.3%; Score 27.8; DB 3; Length 2274;
Best Local Similarity 57.5%; Pred. No. 11;
Matches 50; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 50 AACTCAGGTGTGCACAATTATCAGGACACCCCAAAAGTGGTGAATAGCAT 109
Db 1848 AACACATTATGTAAAAAAGAGACACACAAAAAAGTGGTGAATAGCAT 1789

QY 110 GAGAAGCCGCTTTGATGTTAATAAT 136
Db 1788 GCCATTCTATATTGATGTTAATAAT 1762

RESULT 11
US-08-716-873-27/c
; Sequence 27, Application US/08716873

```

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.4 mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/414,006  
FILING DATE: October 7, 1999  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/492,459  
FILING DATE: June 20, 1995  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: 08/243,403  
FILING DATE: May 16, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren M. Cheek, Jr.  
REGISTRATION NUMBER: 33,367

REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-721-8200  
TELEFAX: 202-721-8250  
TELEX:  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2274  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
US-09-414-006-13

Query Match 20.3%; Score 27.8; DB 4; Length 2274;  
Best Local Similarity 57.5%; Pred. No. 11;  
Matches 50; Conservative 0; Mismatches 37; Indels 0; Gaps 0;  
QY 50 AACTCAGGTGTCACATTTATCAGGACACCCCAACCAAGTGGTAGAATAGCAT 109  
DB 1848 AACACATTAAATGTAAGAAAAAGAAAGACAAACACAAAAAAGAACTTAAAAAACTGT 1789  
QY 110 GAGAACGCGTGTGGTGGTAAATTAAT 136  
DB 1788 GCCATTCTATATTGATGGTAATTAAT 1762

RESULT 14  
US-09-447-223-13/c  
Sequence 13, Application US/09447223  
Patent No. 6432664  
GENERAL INFORMATION:  
APPLICANT: Takashi OKADO et al.  
TITLE OF INVENTION: A GENE CODING FOR A PROTEIN REGULATING  
AUROBASITIN SENSITIVITY

NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wenderoth, Lind & Ponack  
STREET: 805 Fifteenth Street, N.W., #700  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch, 500 kb  
COMPUTER: IBM compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/447,223  
FILING DATE: 23-No. 6432664-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/243,403  
FILING DATE: May 16, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren M. Cheek, Jr.  
REGISTRATION NUMBER: 33,367  
REFERENCE/DOCKET NUMBER: <Unknown>  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-8850  
TELEFAX: <Unknown>  
TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2274  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
SEQUENCE DESCRIPTION: SEQ ID NO: 13:  
US-09-447-223-13

Query Match 20.3%; Score 27.8; DB 4; Length 2274;  
Best Local Similarity 57.5%; Pred. No. 11;  
Matches 50; Conservative 0; Mismatches 37; Indels 0; Gaps 0;  
QY 50 AACTCAGGTGTCACATTTATCAGGACACCCCAACCAAGTGGTAGAATAGCAT 109  
DB 1848 AACACATTAAATGTAAGAAAAAGAAAGACAAACACAAAAAAGAACTTAAAAAACTGT 1789  
QY 110 GAGAACGCGTGTGGTGGTAAATTAAT 136  
DB 1788 GCCATTCTATATTGATGGTAATTAAT 1762

RESULT 15  
US-08-874-186-42  
Sequence 42, Application US/08874186  
Patent No. 5989885  
GENERAL INFORMATION:  
APPLICANT: Teng, David H-F.  
APPLICANT: Tavtigian, Sean V.  
APPLICANT: Perry III, William L.  
APPLICANT: Skolnick, Mark H.  
TITLE OF INVENTION: SPECIFIC MUTATIONS OF MAP KINASE KINASE  
IN HUMAN TUMOR CELL LINES IDENTIFY IT AS A TUMOR  
TITLE OF INVENTION: SUPPRESSOR IN VARIOUS TYPES OF CANCER  
NUMBER OF SEQUENCES: 96  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP  
STREET: 1201 New York Avenue, N.W., Suite 1000  
CITY: Washington  
STATE: DC  
COUNTRY: U.S.A.  
ZIP: 20005

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/874,186  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/782,482  
FILING DATE: 10-JAN-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Saxe, Stephen A.  
REGISTRATION NUMBER: 38,609  
REFERENCE/DOCKET NUMBER: 24884-121392-01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-962-4848  
TELEFAX: 202-962-8300

INFORMATION FOR SEQ ID NO: 42:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2799 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: intron  
LOCATION: 1..764  
FEATURE:  
NAME/KEY: exon  
LOCATION: 765..892  
FEATURE:  
NAME/KEY: intron  
LOCATION: 893..2799  
US-08-874-186-42

Query Match 20.3%; Score 27.8; DB 2; Length 2799;

Best Local Similarity 57.5%; Pred. No. 12;  
Matches 50; Conservative 0; Mismatches 37; Indels 0; Gaps 0;  
Qy 18 AATACCGCAATGGACAACTGGTGAAGCTGTGAACCTCAGGTGTGCACAAATTATCAGGAAC 77  
Db 2082 AATATAAATAAATTAGACATGGAGTATTTGGGAATGTAATTACATATTTTATTAGGAT 2141  
Qy 78 ACCCAAAACCAAGTGAGGTAGAAAT 104  
Db 2142 ACCCGAGAACTGAATGTTAT 2168

Search completed: April 6, 2004, 14:14:17  
Job time : 16.2357 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 15:01:53 ; Search time 61.7891 Seconds  
(without alignments)

8309.593 Million cell updates/sec

Title: US-09-966-264D-1

Perfect score: 137  
Sequence: 1 attataaggagaaagaaaa.....gtgttgatgtaattaatt 137

Scoring table: OLIGO NUC

Gapop 60.0 , Gapext 60.0

Searched: 2470632 seqs, 1873875610 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4941264

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications NA:\*\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US05\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 17: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 18: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	137	100.0	137	9	US-09-966-264-1
C 2	137	100.0	430	9	Sequence 1, Appli
C 3	137	100.0	430	14	Sequence 3505, Ap
C 4	137	100.0	430	15	Sequence 3505, Ap
C 5	137	100.0	430	15	Sequence 3505, Ap
C 6	137	100.0	996	9	Sequence 2, Appli
C 7	137	100.0	2691	15	Sequence 38, Appl
C 8	137	100.0	8689	15	Sequence 42, Appl
C 9	137	100.0	11443	15	Sequence 44, Appl
C 10	137	100.0	13957	9	Sequence 22, Appl
C 11	137	100.0	13957	9	Sequence 2284, Ap
C 12	137	100.0	13957	15	Sequence 1, Appli
C 13	137	100.0	14069	12	Sequence 434, App
C 14	137	100.0	14082	12	Sequence 981, App
C 15	137	100.0	14082	15	Sequence 108, App

C 16	13815	15	US-10-149-736-2	Sequence 2, Appli
C 17	22	22	US-09-966-264-13	Sequence 13, Appl
C 18	20	16.1	US-10-131-827-8211	Sequence 8211, Ap
C 19	19	13.9	US-10-027-632-293205	Sequence 293205,
C 20	19	13.9	US-10-027-632-81394	Sequence 81394, A
C 21	19	13.9	US-10-027-632-81394	Sequence 81394, A
C 22	19	13.9	US-10-425-114-6882	Sequence 6882, Ap
C 23	19	13.9	US-09-373-658-23	Sequence 23, Appl
C 24	19	13.9	US-09-989-687-23	Sequence 23, Appl
C 25	18	13.1	US-10-235-192A-27	Sequence 27, Appl
C 26	18	13.1	US-09-966-264-22	Sequence 22, Appl
C 27	18	13.1	US-10-027-632-274023	Sequence 274023,
C 28	18	13.1	US-10-027-632-280438	Sequence 280438,
C 29	18	13.1	US-10-282-132A-21513	Sequence 21513, A
C 30	18	13.1	US-10-027-632-148232	Sequence 148232,
C 31	18	13.1	US-10-027-632-154166	Sequence 154166,
C 32	18	13.1	US-10-027-632-154167	Sequence 154167,
C 33	18	13.1	US-10-027-632-154168	Sequence 154168,
C 34	18	13.1	US-10-027-632-154169	Sequence 154169,
C 35	18	13.1	US-09-770-445-538	Sequence 538, App
C 36	18	13.1	US-10-027-632-258397	Sequence 258397,
C 37	18	13.1	US-10-027-632-258398	Sequence 258398,
C 38	18	13.1	US-10-027-632-102777	Sequence 102777,
C 39	18	13.1	US-09-814-353-19075	Sequence 19075, A
C 40	18	13.1	US-10-135-322-34	Sequence 34, Appl
C 41	18	13.1	US-10-135-322-33	Sequence 33, Appl
C 42	17	12.4	US-10-135-322-13	Sequence 13, Appl
C 43	17	12.4	US-10-085-783A-12712	Sequence 12712, A
C 44	17	12.4	US-10-242-535A-12712	Sequence 12712, A
C 45	17	12.4	US-10-424-599-396	Sequence 396, App
			US-09-918-995-13211	Sequence 13211, A

## ALIGNMENTS

RESULT 1  
US-09-966-264-1  
; Sequence 1, Application US/09966264  
; Patent No. US2002009015A1  
; GENERAL INFORMATION:  
; APPLICANT: Barber, Elizabeth K  
; TITLE OF INVENTION: Gene Expression Control Element DNA  
; FILE REFERENCE: 8960346US001  
; CURRENT APPLICATION NUMBER: US/09/966,264  
; CURRENT FILING DATE: 2001-09-28  
; PRIOR APPLICATION NUMBER: US 60/237,079  
; PRIOR FILING DATE: 2000-09-30  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 1  
; LENGTH: 137  
; TYPE: DNA  
; ORGANISM: human  
; FEATURE:  
; NAME/KEY: exon  
; LOCATION: (3)...(137)  
; OTHER INFORMATION:  
; NAME/KEY: polyA\_site  
; LOCATION: (130)...(135)  
; OTHER INFORMATION:  
US-09-966-264-1

Query Match	100.0%	Score 137;	DB 9;	Length 137;
Best Local Similarity	100.0%	Pred. No. 9.5e-61;	Mismatches 0;	Indels 0;
Matches 137;	Conservative	0;		
Qy	1	ATTATAAGGAAAAAATAACGCAATGCAAGTGTGAAGCTGTGAACCTCAGTGT	60	
Db	1	ATTATAAGGAAAAAATAACGCAATGCAAGTGTGAAGCTGTGAACCTCAGTGT	60	
Qy	61	GCACAAATATCAGGAACACCCCAACCAAGTAGAGTAAGTAAGTACGAGAGCCGCTG	120	
Db	61	GCACAAATATCAGGAACACCCCAACCAAGTAGAGTAAGTAAGTACGAGAGCCGCTG	120	

QY 121 TTTGATGTTAAATAATT 137  
 |||||  
 Db 121 TTTGATGTTAAATAATT 137

RESULT 2

US-09-796-692-3505/c  
 ; Sequence 3505, Application US/09796692  
 ; Publication No. US20020198362A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gaiger, Alexander  
 ; APPLICANT: Algate, Paul A.  
 ; APPLICANT: Mannion, Jane

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY  
 ; TITLE OF INVENTION: Hematological Malignancies

; FILE REFERENCE: 2077.001200  
 ; CURRENT APPLICATION NUMBER: US/09/796,692

; PRIOR FILING DATE: 2001-03-01  
 ; PRIOR APPLICATION NUMBER: 60/186,126  
 ; PRIOR FILING DATE: 2000-03-01  
 ; PRIOR APPLICATION NUMBER: 60/190,479  
 ; PRIOR FILING DATE: 2000-03-17  
 ; PRIOR APPLICATION NUMBER: 60/200,545  
 ; PRIOR FILING DATE: 2000-04-27  
 ; PRIOR APPLICATION NUMBER: 60/200,303  
 ; PRIOR FILING DATE: 2000-04-28  
 ; PRIOR APPLICATION NUMBER: 60/200,779  
 ; PRIOR FILING DATE: 2000-04-28  
 ; PRIOR APPLICATION NUMBER: 60/200,999  
 ; PRIOR FILING DATE: 2000-05-01  
 ; PRIOR APPLICATION NUMBER: 60/202,084  
 ; PRIOR FILING DATE: 2000-05-04  
 ; PRIOR APPLICATION NUMBER: 60/206,201  
 ; PRIOR FILING DATE: 2000-05-22  
 ; PRIOR APPLICATION NUMBER: 60/218,950  
 ; PRIOR FILING DATE: 2000-07-14  
 ; PRIOR APPLICATION NUMBER: 60/222,903  
 ; PRIOR FILING DATE: 2000-08-03  
 ; PRIOR APPLICATION NUMBER: 60/223,416  
 ; PRIOR FILING DATE: 2000-08-04  
 ; PRIOR APPLICATION NUMBER: 60/223,378  
 ; PRIOR FILING DATE: 2000-08-07  
 ; PRIOR APPLICATION NUMBER: 60/223,692  
 ; NUMBER OF SEQ ID NOS: 9597  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 3505  
 ; LENGTH: 430  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; ORGANISM: Homo sapiens  
 ; US-09-796-692-3505

Query Match 100.0%; Score 137; DB 9; Length 430;

Best Local Similarity 100.0%; Pred. No. 1e-60;

Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATAAGGAAAAAGAAATAACGCAATGGCAAGTGGTGAAGCTGTGAAGCTGAACTCAGGTGT 60  
 |||||

Db 287 ATTATAAGGAAAAAGAAATAACGCAATGGCAAGTGGTGAAGCTGTGAAGCTGAACTCAGGTGT 228  
 |||||

QY 61 GCACAATTATCAGGAACACCCCAAAACCAAGTAGGAGTAAATAGCATGAGAGCCGTG 120  
 |||||

Db 227 GCACATTATCAGGAACACCCCAAAACCAAGTAGGAGTAAATAGCATGAGAGCCGTG 168  
 |||||

QY 121 TTTGATGTTAAATAATT 137  
 |||||

Db 167 TTTGATGTTAAATAATT 151  
 |||||

RESULT 3

US-10-040-862-3505/c  
 ; Sequence 3505, Application US/10040862  
 ; Publication No. US20030078396A1  
 ; GENERAL INFORMATION:

; APPLICANT: Gaiger, Alexander  
 ; APPLICANT: Algate, Paul A.  
 ; APPLICANT: Mannion, Jane  
 ; APPLICANT: Retter, Marc  
 ; APPLICANT: Corixa Corporation  
 ; TITLE OF INVENTION: Hematological Malignancies  
 ; TITLE OF INVENTION: Hematological Malignancies  
 ; FILE REFERENCE: 014058-013520US  
 ; CURRENT APPLICATION NUMBER: US/10/040,862  
 ; CURRENT FILING DATE: 2001-11-06  
 ; PRIOR APPLICATION NUMBER: US 60/186,126  
 ; PRIOR FILING DATE: 2000-03-01  
 ; PRIOR APPLICATION NUMBER: US 60/190,479  
 ; PRIOR FILING DATE: 2000-03-17  
 ; PRIOR APPLICATION NUMBER: US 60/200,545  
 ; PRIOR FILING DATE: 2000-04-27  
 ; PRIOR APPLICATION NUMBER: US 60/200,303  
 ; PRIOR FILING DATE: 2000-04-28  
 ; PRIOR APPLICATION NUMBER: US 60/200,779  
 ; PRIOR FILING DATE: 2000-04-28  
 ; PRIOR APPLICATION NUMBER: US 60/200,999  
 ; PRIOR FILING DATE: 2000-05-01  
 ; PRIOR APPLICATION NUMBER: US 60/202,084  
 ; PRIOR FILING DATE: 2000-05-04  
 ; PRIOR APPLICATION NUMBER: US 60/206,201  
 ; PRIOR FILING DATE: 2000-05-22  
 ; PRIOR APPLICATION NUMBER: US 60/218,950  
 ; PRIOR FILING DATE: 2000-07-14  
 ; PRIOR APPLICATION NUMBER: US 60/222,903  
 ; PRIOR FILING DATE: 2000-08-03  
 ; PRIOR APPLICATION NUMBER: US 60/223,416  
 ; PRIOR FILING DATE: 2000-08-04  
 ; PRIOR APPLICATION NUMBER: US 60/223,378  
 ; PRIOR FILING DATE: 2000-08-07  
 ; PRIOR APPLICATION NUMBER: US 09/796,692  
 ; NUMBER OF SEQ ID NOS: 10467  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 3505  
 ; LENGTH: 430  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; ORGANISM: Homo sapiens  
 ; US-10-040-862-3505

Query Match 100.0%; Score 137; DB 14; Length 430;

Best Local Similarity 100.0%; Pred. No. 1e-60;

Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATAAGGAAAAAGAAATAACGCAATGGCAAGTGGTGAAGCTGTGAAGCTGAACTCAGGTGT 60  
 |||||

Db 287 ATTATAAGGAAAAAGAAATAACGCAATGGCAAGTGGTGAAGCTGTGAAGCTGAACTCAGGTGT 228  
 |||||

QY 61 GCACAATTATCAGGAACACCCCAAAACCAAGTAGGAGTAAATAGCATGAGAGCCGTG 120  
 |||||

Db 227 GCACATTATCAGGAACACCCCAAAACCAAGTAGGAGTAAATAGCATGAGAGCCGTG 168  
 |||||

QY 121 TTTGATGTTAAATAATT 137  
 |||||

Db 167 TTTGATGTTAAATAATT 151  
 |||||

RESULT 4

US-10-057-475B-3505/c  
 ; Sequence 3505, Application US/10057475B  
 ; Publication No. US2004002068A1  
 ; GENERAL INFORMATION:

; APPLICANT: Gaiger, Alexander  
 ; APPLICANT: Algate, Paul A.  
 ; APPLICANT: Mannion, Jane  
 ; APPLICANT: Clapper, Jonathan David  
 ; APPLICANT: Wang, Aijun  
 ; APPLICANT: Ordonez, Nadia  
 ; APPLICANT: Carter, Lauren

```

; APPLICANT: McNeill, Patricia Dianne
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; TITLE OF INVENTION: Hematological Malignancies
; FILE REFERENCE: 014058-014402US
; CURRENT APPLICATION NUMBER: US/10/057,475B
; CURRENT FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1079
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3505
; LENGTH: 430
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-057-475B-3505

Query Match      100.0%; Score 137; DB 15; Length 430;
Best Local Similarity 100.0%; Pred. No. 1e-60;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATAAGGAAAAAGAAATAACCAATGCAAGTGGTGAAGCTGTGAAGCTGTGAAGCTGTGAAGCTGT 60
DB 287 ATTATAAGGAAAAAGAAATAACCAATGCAAGTGGTGAAGCTGTGAAGCTGTGAAGCTGT 228
QY 61 GCACAATTATCAGGAACACCCCAAAACCAAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 120
DB 227 GCACAATTATCAGGAACACCCCAAAACCAAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 168
QY 121 TTGTGATGTTAATTAATT 137
DB 167 TTGTGATGTTAATTAATT 151

RESULT 5
US-10-154-884B-3505/c
; Sequence 3505, Application US/10154884B
; Publication No. US20040005561A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc W.
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; TITLE OF INVENTION: Hematological Malignancies
; FILE REFERENCE: 014058-013521US
; CURRENT APPLICATION NUMBER: US/10/154,884B
; CURRENT FILING DATE: 2002-05-23
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545

```

```

; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 11290
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3505
; LENGTH: 430
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-154-884B-3505

Query Match      100.0%; Score 137; DB 15; Length 430;
Best Local Similarity 100.0%; Pred. No. 1e-60;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATAAGGAAAAAGAAATAACCAATGCAAGTGGTGAAGCTGTGAAGCTGTGAAGCTGT 60
DB 287 ATTATAAGGAAAAAGAAATAACCAATGCAAGTGGTGAAGCTGTGAAGCTGTGAAGCTGT 228
QY 61 GCACAATTATCAGGAACACCCCAAAACCAAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 120
DB 227 GCACAATTATCAGGAACACCCCAAAACCAAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 168
QY 121 TTGTGATGTTAATTAATT 137
DB 167 TTGTGATGTTAATTAATT 151

RESULT 6
US-09-966-264-2
; Sequence 2, Application US/09966264
; Patent No. US2002009015A1
; GENERAL INFORMATION:
; APPLICANT: Barber, Elizabeth K
; TITLE OF INVENTION: Gene Expression Control Element DNA
; FILE REFERENCE: 996034605001
; CURRENT APPLICATION NUMBER: US/09/966,264
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US 60/237,079
; PRIOR FILING DATE: 2000-09-30
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 996
; TYPE: DNA
; ORGANISM: human
; FEATURE:
; NAME/KEY: exon
; LOCATION: (1)..(996)
; OTHER INFORMATION:
; NAME/KEY: misc feature
; LOCATION: (710)..(996)
; OTHER INFORMATION: Nucleotides 710-996 are homologous to a portion of human dystroph
; OTHER INFORMATION: in DNA in the region of exon 79 except that nucleotides 860-996 a
; OTHER INFORMATION: re inverted in comparison to the orientation of the same sequence
; OTHER INFORMATION: in the dystrophin DNA
US-09-966-264-2

Query Match      100.0%; Score 137; DB 9; Length 996;
Best Local Similarity 100.0%; Pred. No. 1e-60;

```





```

; APPLICANT: Sandalon, Ziv.
; APPLICANT: Gnatenko, Dmitri
; TITLE OF INVENTION: Adenoviral Vectors
; FILE REFERENCE: STONYB-04970
; CURRENT APPLICATION NUMBER: US/09/782,378A
; CURRENT FILING DATE: 2001-02-12
; PRIOR APPLICATION NUMBER: 60/237,747
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 13957
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-782-378A-22

Query Match      100.0%; Score 137; DB 9; Length 13957;
Best Local Similarity 100.0%; Pred. No. 1.2e-60;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATAAGGAAAAAGAAAAATACCGAATGGCAAGTGGTGAAGCTGTGAACCTCAGGTGT 60
DB 13299 ATTATAAGGAAAAAGAAAAATACCGAATGGCAAGTGGTGAAGCTGTGAACCTCAGGTGT 13240
QY 61 GCACAAATTATCAGGAACACCCCAAAACCAAACTGAGGTAGAAATAGCATGAGAACCCGTG 120
DB 13239 GCACAAATTATCAGGAACACCCCAAAACCAAACTGAGGTAGAAATAGCATGAGAACCCGTG 13180
QY 121 TTGTGATGTTAATAATT 137
DB 13179 TTGTGATGTTAATAATT 13163

RESULT 11
US-09-880-107-2284/c
; Sequence 2284, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-S028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2284
; LENGTH: 13957
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 M18533
US-09-880-107-2284

Query Match      100.0%; Score 137; DB 9; Length 13957;
Best Local Similarity 100.0%; Pred. No. 1.2e-60;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATAAGGAAAAAGAAAAATACCGAATGGCAAGTGGTGAAGCTGTGAACCTCAGGTGT 60
DB 13299 ATTATAAGGAAAAAGAAAAATACCGAATGGCAAGTGGTGAAGCTGTGAACCTCAGGTGT 13240
QY 61 GCACAAATTATCAGGAACACCCCAAAACCAAACTGAGGTAGAAATAGCATGAGAACCCGTG 120
DB 13239 GCACAAATTATCAGGAACACCCCAAAACCAAACTGAGGTAGAAATAGCATGAGAACCCGTG 13180
QY 121 TTGTGATGTTAATAATT 137
DB 13179 TTGTGATGTTAATAATT 13163

RESULT 12
US-10-149-736-1/c
; Sequence 1, Application US/10149736
; Publication No. US20030216332A1
; GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; APPLICANT: Harper, Scott O.
; TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences
; FILE REFERENCE: UM-06968
; CURRENT APPLICATION NUMBER: US/10/149,736
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: PCT/US01/31126
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,848
; PRIOR FILING DATE: 2000-10-06
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 13957
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-149-736-1

Query Match      100.0%; Score 137; DB 15; Length 13957;
Best Local Similarity 100.0%; Pred. No. 1.2e-60;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATAAGGAAAAAGAAAAATACCGAATGGCAAGTGGCAAGTGGTGAAGCTGTGAACCTCAGGTGT 60
DB 13299 ATTATAAGGAAAAAGAAAAATACCGAATGGCAAGTGGCAAGTGGTGAAGCTGTGAACCTCAGGTGT 13240
QY 61 GCACAAATTATCAGGAACACCCCAAAACCAAACTGAGGTAGAAATAGCATGAGAACCCGTG 120
DB 13239 GCACAAATTATCAGGAACACCCCAAAACCAAACTGAGGTAGAAATAGCATGAGAACCCGTG 13180
QY 121 TTGTGATGTTAATAATT 137
DB 13179 TTGTGATGTTAATAATT 13163

RESULT 13
US-10-342-887-434/c
; Sequence 434, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 434
; LENGTH: 14069
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-434
```

Query Match 100.0%; Score 137; DB 12; Length 14069;  
Best Local Similarity 100.0%; Pred. No. 1.2e-60;  
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATTATAAGGAAAAAGAAAAATAACCAATGGACAAGTGTGAAGCTGTGAAGTGTGAAGTGTGT 60  
Db 13411 ATTATAAGGAAAAAGAAAAATAACCAATGGACAAGTGTGAAGCTGTGAAGTGTGT 13352  
QY 61 GCACAATTATCAGGAACACCCCAAAACCAAAAGTGAAGTGTGAAGTGTGAAGTGTGAAGTGT 120  
Db 13351 GCACAATTATCAGGAACACCCCAAAACCAAAAGTGAAGTGTGAAGTGTGAAGTGTGAAGTGT 13292  
QY 121 TTTGATGTTAATTAATT 137  
Db 13291 TTTGATGTTAATTAATT 13275

RESULT 14  
US-10-342-887-981/c  
; Publication 981, Application US/10342887  
; Sequence 981, Application US/10342887  
; Publication No. US20040058340A1  
; GENERAL INFORMATION:  
; APPLICANT: Dai, Hongyue  
; APPLICANT: He, Yudong  
; APPLICANT: Linsley, Peter S.  
; APPLICANT: Mao, Mao  
; APPLICANT: Roberts, Christopher J.  
; APPLICANT: Van 't Veer, Laura Johanna  
; APPLICANT: Van de Vijver, Marc J.  
; APPLICANT: Bernards, Rene  
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients  
; FILE REFERENCE: 9301-188-999  
; CURRENT APPLICATION NUMBER: US/10/342,887  
; CURRENT FILING DATE: 2003-01-15  
; PRIOR APPLICATION NUMBER: 60/298,918  
; PRIOR FILING DATE: 2001-06-18  
; PRIOR APPLICATION NUMBER: 60/380,710  
; PRIOR FILING DATE: 2002-05-14  
; PRIOR APPLICATION NUMBER: 10/172,118  
; PRIOR FILING DATE: 2002-06-14  
; NUMBER OF SEQ ID NOS: 2699  
; SEQ ID NO 981  
; LENGTH: 14082  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-342-887-981

Query Match 100.0%; Score 137; DB 12; Length 14082;  
Best Local Similarity 100.0%; Pred. No. 1.2e-60;  
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATTATAAGGAAAAAGAAAAATAACCAATGGACAAGTGTGAAGCTGTGAAGTGTGAAGTGTGT 60  
Db 13424 ATTATAAGGAAAAAGAAAAATAACCAATGGACAAGTGTGAAGCTGTGAAGTGTGT 13365  
QY 61 GCACAATTATCAGGAACACCCCAAAACCAAAAGTGAAGTGTGAAGTGTGAAGTGTGAAGTGT 120  
Db 13364 GCACAATTATCAGGAACACCCCAAAACCAAAAGTGAAGTGTGAAGTGTGAAGTGTGAAGTGT 13305  
QY 121 TTTGATGTTAATTAATT 137  
Db 13304 TTTGATGTTAATTAATT 13288

RESULT 15  
US-10-341-434-108/c  
; Sequence 108, Application US/10341434  
; Publication No. US20030215835A1  
; GENERAL INFORMATION:  
; APPLICANT: Origene Technologies  
; TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes  
; FILE REFERENCE: 90 204 205 R1

; CURRENT APPLICATION NUMBER: US/10/341,434  
; CURRENT FILING DATE: 2003-07-18  
; PRIOR APPLICATION NUMBER: US 60/348,164  
; PRIOR FILING DATE: 2002-01-15  
; PRIOR APPLICATION NUMBER: US 60/348,119  
; PRIOR FILING DATE: 2002-01-15  
; NUMBER OF SEQ ID NOS: 238  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 108  
; LENGTH: 14082  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (703)..(11388)  
; OTHER INFORMATION:  
US-10-341-434-108

Query Match 100.0%; Score 137; DB 15; Length 14082;  
Best Local Similarity 100.0%; Pred. No. 1.2e-60;  
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATTATAAGGAAAAAGAAAAATAACCAATGGACAAGTGTGAAGCTGTGAAGTGTGAAGTGTGT 60  
Db 13424 ATTATAAGGAAAAAGAAAAATAACCAATGGACAAGTGTGAAGCTGTGAAGTGTGT 13365  
QY 61 GCACAATTATCAGGAACACCCCAAAACCAAAAGTGAAGTGTGAAGTGTGAAGTGTGAAGTGT 120  
Db 13364 GCACAATTATCAGGAACACCCCAAAACCAAAAGTGAAGTGTGAAGTGTGAAGTGTGAAGTGT 13305  
QY 121 TTTGATGTTAATTAATT 137  
Db 13304 TTTGATGTTAATTAATT 13288

Search completed: April 6, 2004, 17:51:12  
Job time : 63.7891 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 14:14:32 ; Search time 15.1147 Seconds  
(without alignments)  
5030.075 Million cell updates/sec

Title: US-09-966-264D-1  
Perfect score: 137  
Sequence: 1 attataaaggaaagaaaa.....gtgtttgtgtaattaatt 137

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA.\*  
1: /cgn2\_6/prodata/2/ina/5A-COMB.seq.\*  
2: /cgn2\_6/prodata/2/ina/5B-COMB.seq.\*  
3: /cgn2\_6/prodata/2/ina/6A-COMB.seq.\*  
4: /cgn2\_6/prodata/2/ina/6B-COMB.seq.\*  
5: /cgn2\_6/prodata/2/ina/PCUS-COMB.seq.\*  
6: /cgn2\_6/prodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	117	85.4	13977	4	US-09-484-970B-60
C 2	42	30.7	19307	3	US-08-836-022A-10
C 3	42	30.7	19307	3	US-09-427-048A-10
C 4	19	13.9	4766	5	PCR-US93-07261-10
C 5	18	13.1	750	4	US-09-107-532A-866
C 6	18	13.1	889	2	US-08-935-886-7
C 7	18	13.1	889	2	US-08-935-886-13
C 8	18	13.1	2076	4	US-09-134-001C-1838
C 9	17	12.4	92	1	US-07-839-751-1
C 10	17	12.4	92	1	US-08-239-427A-1
C 11	17	12.4	92	2	US-08-467-975-1
C 12	17	12.4	92	3	US-08-401-355-1
C 13	17	12.4	1260	3	US-09-081-689-3
C 14	17	12.4	1297	4	US-08-858-207A-66
C 15	17	12.4	1329	4	US-09-305-984-13
C 16	17	12.4	1329	4	US-09-073-541A-13
C 17	17	12.4	1329	4	US-09-493-940-13
C 18	17	12.4	1526	3	US-09-081-689-1
C 19	17	12.4	2817	5	PCR-US93-05944-1
C 20	17	12.4	3356	3	US-09-379-523-4
C 21	17	12.4	3808	2	US-08-417-210A-79
C 22	17	12.4	3808	4	US-09-136-159A-79
C 23	17	12.4	4358	4	US-09-308-453-1
C 24	17	12.4	5152	4	US-10-204-708-48
C 25	17	12.4	5815	4	US-10-204-708-77
C 26	17	12.4	7816	3	US-08-815-809-6
C 27	17	12.4	7797	2	US-08-816-155B-7

## ALIGNMENTS

```

RESULT 1
US-09-484-970B-60/c
; Sequence 60, Application US/09484970B
; Patent No. 6426186
; GENERAL INFORMATION:
; APPLICANT: Jones, Karen A.
; APPLICANT: Volkmut, Wayne
; APPLICANT: Walker, Michael G.
; TITLE OF INVENTION: BONE REMODELING GENES
; FILE REFERENCE: PB-0014 US
; CURRENT APPLICATION NUMBER: US/09/484,970B
; CURRENT FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: PERL Program
; SEQ ID NO 60
; LENGTH: 13977
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6426186 229357.11CBI
; NAME/KEY: unsure
; LOCATION: 11721-11761, 12294, 13969
; OTHER INFORMATION: a, t, c, g, or other
US-09-484-970B-60

Query Match      85.4%   Score 117;   DB 4;   Length 13977;
Best Local Similarity 100.0%;   Pred. No. 3.6e-51;   Mismatches 0;   Indels 0;   Gaps 0;
Matches 117;   Conservative 0;

QY 1 ATTATAAGGAAAGAAATAACGCAATGGCAAGTGTGAAGCTGTGAACCTCAGGTGT 60
DB 13306 ATTATAAGGAAAGAAATAACGCAATGGCAAGTGTGAAGCTGTGAACCTCAGGTGT 13247

QY 61 GCACAAATATCAGGAACACCCGCAACCAAGTGAAGTGAATAGCATGAGAGGCC 117
DB 13246 GCACAAATATCAGGAACACCCGCAACCAAGTGAAGTGAATAGCATGAGAGGCC 13190

RESULT 2
US-08-836-022A-10
; Sequence 10, Application US/08836022A
; Patent No. 6001557
; GENERAL INFORMATION:
; APPLICANT: Trustees of the University of Pennsylvania
; APPLICANT: Wilson, James M.
; APPLICANT: Fisher, Krishna J.
; APPLICANT: Cher, Shu-Jen
; APPLICANT: Weitzman, Matthew
; TITLE OF INVENTION: Improved Adenovirus Virus and

```

```

28 17 12.4 7797 3 US-09-079-587-7 Sequence 7, Appli
C 29 17 12.4 8900 4 US-09-305-984-25 Sequence 25, Appl
C 30 17 12.4 8900 4 US-09-073-541A-25 Sequence 25, Appl
C 31 17 12.4 8900 4 US-09-493-940-25 Sequence 25, Appl
32 17 12.4 10962 2 US-08-816-155B-6 Sequence 6, Appli
33 17 12.4 10962 3 US-09-079-587-6 Sequence 6, Appli
34 17 12.4 19877 2 US-08-816-155B-8 Sequence 8, Appli
35 17 12.4 19877 3 US-09-079-587-8 Sequence 8, Appli
36 17 12.4 28882 4 US-08-361-527-140 Sequence 140, App
C 37 16 11.7 312 4 US-09-543-681A-1844 Sequence 1844, App
38 16 11.7 402 4 US-09-621-976-113 Sequence 113, App
39 16 11.7 439 4 US-09-397-787-294 Sequence 294, App
40 16 11.7 661 4 US-09-702-705-219 Sequence 219, App
41 16 11.7 661 4 US-09-736-457-219 Sequence 219, App
42 16 11.7 661 4 US-09-614-124B-219 Sequence 219, App
43 16 11.7 681 4 US-08-671-325-219 Sequence 219, App
44 16 11.7 681 4 US-09-589-184-219 Sequence 219, App
45 16 11.7 1247 4 US-09-740-035-1 Sequence 1, Appli

```

NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Howson and Howson  
STREET: Spring House Corporate Cntr, P O Box 457  
CITY: Spring House  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19477  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/836,022A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/331,381  
FILING DATE: 28-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: GNVEN.008PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-540-9200  
TELEFAX: 215-540-5818  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19307 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: cDNA  
US-08-836-022A-10

Query Match 30.7%; Score 42; DB 3; Length 19307;  
Best Local Similarity 100.0%; Pred. No. 1.5e-12; Mismatches 0; Indels 0; Gaps 0;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 AACACCCCAAAACCAAGTGAGGTAGAGTAAATAGCATGAGAAGC 116  
|||||  
DB 1563 AACACCCCAAAACCAAGTGAGGTAGAGTAAATAGCATGAGAAGC 1604

RESULT 3  
US-09-427-048A-10  
Sequence 10, Application US/09427048A  
Patent No. 6203975  
GENERAL INFORMATION:  
APPLICANT: Trustees of the University of Pennsylvania  
Wilson, James M.  
Fisher, Krishna J.  
Chen, Shu-Jen  
Weitzman, Matthew  
TITLE OF INVENTION: Improved Adenovirus Virus and Methods of Use Thereof  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Howson and Howson  
STREET: Spring House Corporate Cntr, P O Box 457  
CITY: Spring House  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19477  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/427,048A  
FILING DATE: 21-Oct-1999

CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/836,022  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: GNVEN.008PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-540-9200  
TELEFAX: 215-540-5818  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19307 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: cDNA  
SEQUENCE DESCRIPTION: SEQ ID NO: 10:  
US-09-427-048A-10

Query Match 30.7%; Score 42; DB 3; Length 19307;  
Best Local Similarity 100.0%; Pred. No. 1.5e-12; Mismatches 0; Indels 0; Gaps 0;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 AACACCCCAAAACCAAGTGAGGTAGAGTAAATAGCATGAGAAGC 116  
|||||  
DB 1563 AACACCCCAAAACCAAGTGAGGTAGAGTAAATAGCATGAGAAGC 1604

RESULT 4  
PCT-US93-07261-10  
Sequence 10, Application PC/TUS9307261  
GENERAL INFORMATION:  
TITLE OF INVENTION: PIEMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: John H. C. Blasdale  
STREET: One Giralda Farms  
CITY: Madison  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07940-1000  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh 6.0.5  
SOFTWARE: Microsoft Word 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/07261  
FILING DATE: 19930805  
PRIOR APPLICATION DATA: US 07/927,531  
APPLICATION NUMBER:  
FILING DATE: 07-AUG-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Blasdale, John H. C.  
REGISTRATION NUMBER: 31,895  
REFERENCE/DOCKET NUMBER: DX0288K  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-822-7398  
TELEFAX: 201-822-7039  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4766 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
ORIGINAL SOURCE:  
ORGANISM: Plasmodium falciparum  
STRAIN: Malayan Camp  
IMMEDIATE SOURCE:  
CLONE: p2b1/p12-1

FEATURE:  
NAME/KEY: CDS  
LOCATION: 3..4766  
PCT-US93-07261-10

Query Match 13.9%; Score 19; DB 5; Length 4766;  
Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 TAAGGAAAGAAATAA 23  
|||||  
DB 190 TAAGGAAAGAAATAA 208  
|||||

RESULT 5  
US-09-107-532A-866  
Sequence 866, Application US/09107532A  
Patent No. 6583275  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02154

COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Deneka  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 866:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 750 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...750  
SEQUENCE DESCRIPTION: SEQ ID NO: 866:

US-09-107-532A-866

Query Match 13.1%; Score 18; DB 4; Length 750;  
Best Local Similarity 100.0%; Pred. No. 3.4;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 CAATGGACAAGTGGTCAA 43  
|||||

Db 117 CAATGGACAAGTGGTCAA 134

RESULT 6  
US-08-935-886-7  
Sequence 7, Application US/08935886  
Patent No. 5994625  
GENERAL INFORMATION:  
APPLICANT: Melchers, Leo S.  
APPLICANT: Sela-Buuriage, Marianne B.  
APPLICANT: Bres-Vloemans, Alexandra A.  
APPLICANT: Ponstein, Anne S.  
APPLICANT: Apotheker-de Groot, Marion  
APPLICANT: Cornelissen, Bernardus J. C.  
TITLE OF INVENTION: Antifungal Chitin Binding Proteins and DNA  
TITLE OF INVENTION: Coding Therefor  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ladass & Parry  
STREET: 26 West 61st Street  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10023

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 14.4 MB storage  
COMPUTER: IBM PC/XT/AT or compatibles  
OPERATING SYSTEM: DOS 6.20  
SOFTWARE: Wordperfect 5.1

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/935,886  
FILING DATE:  
CLASSIFICATION: 800

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/411,640  
FILING DATE: 05-APR-1995  
APPLICATION NUMBER: PCT/EP93/02790  
FILING DATE: 05-OCT-1993

ATTORNEY/AGENT INFORMATION:  
NAME: Mass, Clifford J.  
REGISTRATION NUMBER: 30,086  
REFERENCE/DOCKET NUMBER: U-010139-3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 708-1800  
TELEFAX: (212) 246-8959  
TELEX: 233288

INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 889 Base Pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
MOLECULE TYPE: cDNA to mRNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Nicotiana tabacum  
STRAIN: Samsun NN  
DEVELOPMENTAL STAGE: Mature  
TISSUE TYPE: Leaf, wounded  
IMMEDIATE SOURCE:  
LIBRARY: lambda zap  
CLONE: CBP4.4  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 16..639  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 1..14  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 866..889  
OTHER INFORMATION: /function= "XhoI-linker"

US-08-935-886-7

Query Match 13.1%; Score 18; DB 2; Length 889;  
Best Local Similarity 100.0%; Pred. No. 3.4;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 AAAGGAAAAGAAATAA 23  
|||||  
Db 707 AAAGGAAAAGAAATAA 724

## RESULT 7

US-08-935-886-13  
; Sequence 13, Application US/08935886  
; Patent No. 5934525  
; GENERAL INFORMATION:  
; APPLICANT: Melchers, Leo S.  
; APPLICANT: Sela-Buurlage, Marianne B.  
; APPLICANT: Bres-Vloemans, Alexandra A.  
; APPLICANT: Ponstein, Anne S.  
; APPLICANT: Apotheke-de Groot, Marion  
; APPLICANT: Cornelissen, Bernardus J. C.  
; TITLE OF INVENTION: Antifungal Chitin Binding Proteins and DNA  
; TITLE OF INVENTION: Coding Therefor  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Ladass & Parry  
; STREET: 26 West 61st Street  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10023

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 14.4 MB storage  
COMPUTER: IBM PC/XT/AT or compatibles  
OPERATING SYSTEM: DOS 6.20  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/935,886  
FILING DATE:

CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/411,640  
FILING DATE: 05-APR-1995  
APPLICATION NUMBER: PCT/EP93/02790  
FILING DATE: 05-OCT-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Mass, Clifford J.  
REGISTRATION NUMBER: 30,086  
REFERENCE/DOCKET NUMBER: U-010139-3  
TELEPHONE: (212) 708-1800  
TELEFAX: (212) 246-8959  
TELEX: 233288  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 889 Base Pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: double  
TOPOLOGY: Linear  
MOLECULE TYPE: cDNA to mRNA  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Nicotiana tabacum  
STRAIN: Samsun NN  
TISSUE TYPE: Leaf, wounded  
IMMEDIATE SOURCE:  
CLONE: CP4.4T  
US-08-935-886-13

Query Match 13.1%; Score 18; DB 2; Length 889;  
Best Local Similarity 100.0%; Pred. No. 3.4;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 6 AAAGGAAAAGAAATAA 23  
|||||  
Db 726 AAAGGAAAAGAAATAA 743

## RESULT 8

US-09-134-001C-1838  
; Sequence 1838, Application US/09134001C  
; Patent No. 8380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 1838  
; LENGTH: 2076  
; TYPE: DNA  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-1838

Query Match 13.1%; Score 18; DB 4; Length 2076;  
Best Local Similarity 100.0%; Pred. No. 3.4;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TTATAAGGAAAAGAAA 19  
|||||  
Db 1374 TTATAAGGAAAAGAAA 1391

## RESULT 9

US-07-839-751-1/c  
; Sequence 1, Application US/07839751  
; Patent No. 5439809  
; GENERAL INFORMATION:  
; APPLICANT: Haynes, Joel  
; APPLICANT: Klein, Michel H  
; TITLE OF INVENTION: Production of Genetically-Engineered  
; TITLE OF INVENTION: Vaccines for Aids and Retroviral Diseases  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA: US/07/839,751  
APPLICATION NUMBER: US/07/839,751  
FILING DATE: 19920615  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-267 MLS:JC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
TELEX: 065-24567 SIMBAS  
INFORMATION FOR SEQ ID NO: 1:

## ; SEQUENCE CHARACTERISTICS:

LENGTH: 92 base pairs  
 TYPE: NUCLEIC ACID  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US-07-839-751-1

Query Match 12.4%; Score 17; DB 1; Length 92;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AGGAAAAAGAAATAAC 24  
 |||||  
 Db 73 AGGAAAAAGAAATAAC 57

## RESULT 10

US-08-239-427A-1/c  
 ; Sequence 1, Application US/08239427A  
 ; Patent No. 5571712

## ; GENERAL INFORMATION:

APPLICANT: Haynes, Joel  
 APPLICANT: Klein, Michel H  
 APPLICANT: Rovinski, Benjamin  
 APPLICANT: Cao, Shi Xian  
 TITLE OF INVENTION: Production of Genetically-Engineered  
 TITLE OF INVENTION: Vaccines for Aids and Retroviral Diseases  
 NUMBER OF SEQUENCES: 2  
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Sim & McBurney  
 STREET: Suite 701, 330 University Avenue  
 CITY: Toronto  
 STATE: Ontario  
 COUNTRY: Canada  
 ZIP: M5G 1R7

## ; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/239,427A  
 FILING DATE: 06-MAY-1994

## ; CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 839,751  
 FILING DATE: 12-OCT-1990

## ; ATTORNEY/AGENT INFORMATION:

NAME: Stewart, Michael I  
 REGISTRATION NUMBER: 24,973  
 REFERENCE/DOCKET NUMBER: 1038-332 MIS:jb  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (416) 595-1155  
 TELEFAX: (416) 595-1163

## ; INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 92 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US-08-239-427A-1

Query Match 12.4%; Score 17; DB 1; Length 92;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AGGAAAAAGAAATAAC 24  
 |||||  
 Db 73 AGGAAAAAGAAATAAC 57

## RESULT 11

US-08-467-975-1/c  
 ; Sequence 1, Application US/08467975  
 ; Patent No. 5985641

## ; GENERAL INFORMATION:

APPLICANT: Haynes, Joel  
 APPLICANT: Klein, Michel H  
 APPLICANT: Rovinski, Benjamin  
 APPLICANT: Cao, Shi Xian  
 TITLE OF INVENTION: Production of Genetically-Engineered  
 TITLE OF INVENTION: Vaccines for Aids and Other Retroviral Diseases  
 NUMBER OF SEQUENCES: 2  
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Sim & McBurney  
 STREET: Suite 701, 330 University Avenue  
 CITY: Toronto  
 STATE: Ontario  
 COUNTRY: Canada  
 ZIP: M5G 1R7

## ; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/467,975  
 FILING DATE: 06-JUN-1995

## ; CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/239,427  
 FILING DATE: 06-MAY-1994

## ; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/839,751  
 FILING DATE: 12-OCT-1990

## ; PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 8912123.7  
 FILING DATE: 13-OCT-1989

## ; ATTORNEY/AGENT INFORMATION:

NAME: Stewart, Michael I  
 REGISTRATION NUMBER: 24,973  
 REFERENCE/DOCKET NUMBER: 1038-456 MIS:vg  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (416) 595-1155  
 TELEFAX: (416) 595-1163

## ; INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 92 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US-08-467-975-1

Query Match 12.4%; Score 17; DB 2; Length 92;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AGGAAAAAGAAATAAC 24  
 |||||  
 Db 73 AGGAAAAAGAAATAAC 57

## RESULT 12

US-08-401-355-1/c  
 ; Sequence 1, Application US/08401355  
 ; Patent No. 6291227

## ; GENERAL INFORMATION:

APPLICANT: Haynes, Joel  
 APPLICANT: Klein, Michel H  
 TITLE OF INVENTION: Production of Genetically-Engineered  
 TITLE OF INVENTION: Vaccines for Aids and Retroviral Diseases  
 NUMBER OF SEQUENCES: 2

;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Sim & McBurney  
;; STREET: Suite 701, 330 University Avenue  
;; CITY: Toronto  
;; STATE: Ontario  
;; COUNTRY: Canada  
;; ZIP: M5R 1R7  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent In Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/401,355  
;; FILING DATE:  
;; CLASSIFICATION: 424  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/839,751  
;; FILING DATE: 15-JUN-1992  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Stewart, Michael I  
;; REGISTRATION NUMBER: 24,973  
;; REFERENCE/DOCKET NUMBER: 1038-424 MIS:as  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (416) 595-1155  
;; TELEFAX: (416) 595-1163  
;; TELEX: 065-24567 SIMBAS  
;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 92 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; US-08-401-355-1

Query Match 12.4%; Score 17; DB 3; Length 92;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AGGAAAAGAAATAAC 24  
DB 73 AGGAAAAGAAATAAC 57

RESULT 13  
US-09-081-689-3/c  
; Sequence 3, Application US/09081689  
; Patent No. 6165992  
; GENERAL INFORMATION:  
; APPLICANT: Wallis, Nicola G.  
; APPLICANT: Zalacain, Magdalenaa  
; APPLICANT: Throup, John  
; APPLICANT: Biswas, Sanjoy  
; TITLE OF INVENTION: Histidine Kinase  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dechert, Price & Rhoads  
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103-2793  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows 95  
; SOFTWARE: FastSeq for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/081,689  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: 60/048,347  
;; FILING DATE: 30-MAY-1997  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Falk, Stephen T  
;; REGISTRATION NUMBER: 36,795  
;; REFERENCE/DOCKET NUMBER: GM10009  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 215-994-2488  
;; TELEFAX: 215-994-2222  
;; TELEX:  
;; INFORMATION FOR SEQ ID NO: 3:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1260 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; US-09-081-889-3

Query Match 12.4%; Score 17; DB 3; Length 1260;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 AAGGAAAAGAAATAA 23  
DB 631 AAGGAAAAGAAATAA 615

RESULT 14  
US-08-858-207A-66/c  
; Sequence 66, Application US/08858207A  
; Patent No. 6348328  
; GENERAL INFORMATION:  
; APPLICANT: Black, Michael  
; APPLICANT: Hodgson, John  
; APPLICANT: Knowles, David  
; APPLICANT: Nicholas, Richard  
; APPLICANT: Stodola, Robert  
; TITLE OF INVENTION: No. 6348328el Compounds  
; NUMBER OF SEQUENCES: 552  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SmithKline Beecham Corporation  
; STREET: 709 Swedeland Road  
; CITY: King of Prussia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19406-0939  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/858,207A  
; FILING DATE: 09-MAY-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/017670  
; FILING DATE: 14-MAY-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gimmi, Edward R  
; REGISTRATION NUMBER: 38,891  
; REFERENCE/DOCKET NUMBER: P50475  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610-270-4478  
; TELEFAX: 610-270-5090  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 66:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1297 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-858-207A-66



Search completed: April 6, 2004, 17:42:27  
Job time : 16.6147 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 11:42:46 ; Search time 64.9329 Seconds  
(without alignments)  
8963.144 Million cell updates/sec

Title: US-09-966-264D-1  
Perfect score: 137  
Sequence: 1 attataaggaagaaagaaa.....gtgtttgatgttaattaatt 137

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0

Searched: 337863 seqs, 2124099041 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N Geneseq 29Jan04: \*  
1: geneseqn1980s: \*  
2: geneseqn1990s: \*  
3: geneseqn2000s: \*  
4: geneseqn2001as: \*  
5: geneseqn2001bs: \*  
6: geneseqn2002as: \*  
7: geneseqn2003as: \*  
8: geneseqn2003bs: \*  
9: geneseqn2003cs: \*  
10: geneseqn2004s: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	137	100.0	137	6	ABK86496 Human Apo
C 2	137	100.0	158	6	ABK86469 Human Apo
C 3	137	100.0	200	6	ABK86468 Human dys
C 4	137	100.0	455	9	AD312504 Human mit
5	137	100.0	996	6	ABK86497 Human Apo
6	137	100.0	1230	6	ABK86462 Human Apo
7	137	100.0	1234	6	ABK86463 Human Apo
C 8	137	100.0	2691	6	ABK81996 Human dys
C 9	137	100.0	8689	6	ABK82000 DNA encod
C 10	137	100.0	11443	6	ABK82002 DNA encod
C 11	137	100.0	13957	6	ABK81959 CDNA enco
C 12	137	100.0	13957	6	ABK10904 Human bre
C 13	137	100.0	13957	6	ABN95786 Gene #228
C 14	137	100.0	13957	6	ABK69900 Human dys
C 15	117	85.4	13977	6	ABK70403 Human bon
C 16	46	33.6	959	6	ABN74601 Bovine em
C 17	42	30.7	13815	2	Aav18885 Mus muscu
C 18	42	30.7	13815	6	ABK81960 CDNA enco
C 19	42	30.7	13815	6	ABK81960 CDNA enco
C 20	42	30.7	13815	6	ABK81960 CDNA enco
C 21	41	29.9	60	6	ABK86471 Human apo
C 22	40	29.2	108	6	ABK86467 Human apo
C 23	22	16.1	22	6	ABK86479 Human apo

C 24	20	14.6	690	6	ABZ08220 Human leu
C 25	19	13.9	716	4	Aak64599 Human imm
26	19	13.9	4766	2	AAQ70102 Malarial
27	19	13.9	7326	3	AAA70095 Plasmodi
C 28	19	13.9	42521	2	Aaz32013 Human MFT
C 29	19	13.9	42521	5	Aac90070 U64857 cD
30	18	13.1	18	6	Abk86487 Human apo
C 31	18	13.1	324	4	Aaz33883 Tobacco p
C 32	18	13.1	407	4	Aak75817 Human imm
C 33	18	13.1	408	4	Aak75816 Human imm
C 34	18	13.1	476	9	ADB51265 Primary r
35	18	13.1	492	7	ABZ76803 Nicotiana
C 36	18	13.1	569	3	AAC55084 Arabidops
37	18	13.1	717	7	ACA33643 Prokaryot
38	18	13.1	750	9	ADC31239 E. faecau
39	18	13.1	878	6	ABN98770 Arabidops
40	18	13.1	889	2	AAQ62459 Chitin bi
41	18	13.1	889	2	AAQ62455 Chitin bi
C 42	18	13.1	896	3	AAC37805 Arabidops
C 43	18	13.1	1028	3	AAC37406 Arabidops
C 44	18	13.1	1155	3	AAC36681 Arabidops
45	18	13.1	2076	6	ABN92375 Staphyloc

## ALIGNMENTS

RESULT 1  
ID ABK86496 standard; DNA; 137 BP.  
XX AC ABK86496;  
XX DT 27-AUG-2002 (first entry)  
XX DE Human Apo-dystrophin-4 inversion sequence.  
XX KW Human; ds; apo-dystrophin-4; inversion sequence; gene therapy;  
XX KW protein truncation; muscular dystrophy; leukaemia.  
XX OS Homo sapiens.  
XX FH Key Location/Qualifiers  
FT CDS 3..137  
FT FT /\*tag= a  
FT FT /product= "Apo-dystrophin-4 peptide appearing as  
FT FT AAU98738"  
FT FT /partial  
FT FT /note= "No start or stop codon shown"  
FT FT /transl\_except= (pos:21..23,aa:Xaa)  
FT FT /transl\_except= (pos:48..50,aa:Xaa)  
FT FT /transl\_except= (pos:93..95,aa:Xaa)  
FT FT /transl\_except= (pos:123..125,aa:Xaa)  
FT FT /transl\_except= (pos:129..131,aa:Xaa)  
FT FT /note= "Xaa= unknown, encoded by in frame stop codon"

GB2368064-A.

24-APR-2002.

16-JAN-2001; 2001GB-00001124.

30-SEP-2000; 2000US-0237079P.

(IMCR ) IMPERIAL CANCER RES TECHNOLOGY LTD.  
(BARB/) BARBER E.

Barber E;

WPI; 2002-429042/46.

P-PSDB; AAU98737.

New human regulatory polynucleotide, useful for treating disorders

PT associated with protein truncation, particularly muscular dystrophy, and  
PT related peptides and antibodies.  
PS Claim 1; Page 169; 222pp; English.  
XX The invention relates to a polynucleotide (I) comprising, or consisting  
CC of, apo-dystrophin-4 inversion sequence appearing as ABK86496, or its  
CC functional equivalents (e.g. the apo-dystrophin-4 cDNA sequence appearing  
CC as ABK86497). Also included are polynucleotides that hybridise to either  
CC strand of (I), a vector containing (I), a cell containing (I) or the  
CC vector, proteins and peptides encoded by (I), a protein homologous with  
CC human dystrophin that is expressed on cell surfaces in vivo antibodies  
CC (Ab) specific for the protein and method of screening for leukemia cells  
CC by analysing DNA for presence of (I) or by detecting presence of (II).  
CC The apo-dystrophin-4 inversion sequence is a regulatory element that  
CC controls expression (transcription and translation) of associated DNA,  
CC and may allow read-through of stop codons. The apo-dystrophin-4 inversion  
CC sequence is used in gene therapy of diseases associated with truncation  
CC of proteins, particularly muscular dystrophy and also leukaemia, but more  
CC generally (I) is a regulatory sequence used to control expression of any  
CC attached gene. Analysis of DNA for (I), or detection of proteins (II)  
CC encoded by (I), can be used to screen for leukaemic cells and related  
CC diseases. Antibodies raised against (II) can be used therapeutically, to  
CC inhibit (II) activity, also to detect (II) in screening assays. The  
CC present sequence is the apo-dystrophin-4 inversion sequence and upstream  
CC genomic region  
XX  
SQ Sequence 137 BP; 56 A; 19 C; 32 G; 30 T; 0 U; 0 Other;  
Query Match 100.0%; Score 137; DB 6; Length 137;  
Best Local Similarity 100.0%; Pred. No. 4.8e-59;  
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATTATAAGGAAAGAAATAACGCAATGCAAGTGGTGAAGCTGTGAAGTCTCAGGTGT 60  
DB 1 ATTATAAGGAAAGAAATAACGCAATGCAAGTGGTGAAGCTGTGAAGTCTCAGGTGT 60  
QY 61 GCACAAATTATCAGGAACACCCCAAAACCAAAAGTGAAGTGAATATGATGAGAGCCGTG 120  
DB 61 GCACAAATTATCAGGAACACCCCAAAACCAAAAGTGAAGTGAATATGATGAGAGCCGTG 120  
QY 121 TTTGATGTTAATTAATT 137  
DB 121 TTTGATGTTAATTAATT 137  
RESULT 2  
ABK86469/c  
ID ABK86469 standard; cDNA; 158 BP.  
XX AC ABK86469;  
XX 27-AUG-2002 (first entry)  
XX Human apo-dystrophin-4 cDNA fragment with inversion breakpoint #2.  
XX Human; ss; apo-dystrophin-4; inversion sequence; gene therapy;  
XX protein truncation; muscular dystrophy; leukaemia; dystrophin.  
XX Homo sapiens.  
XX Key Location/Qualifiers  
XX misc\_recomb 137  
XX /tag= a  
XX /label= Inversion\_breakpoint  
XX GB2368064-A.  
XX 24-APR-2002.  
XX 16-JAN-2001; 2001GB-00001124.  
XX 30-SEP-2000; 2000US-0237079P.  
XX

XX (TMCR ) IMPERIAL CANCER RES TECHNOLOGY LTD.  
PA (BARE//) BARBER E.  
XX Barber E;  
XX WPI; 2002-429042/46.  
XX New human regulatory polynucleotide, useful for treating disorders  
PT associated with protein truncation, particularly muscular dystrophy, and  
PT related peptides and antibodies.  
XX Disclosure; Fig 16B; 222pp; English.  
XX The invention relates to a polynucleotide (I) comprising, or consisting  
CC of, apo-dystrophin-4 inversion sequence appearing as ABK86496, or its  
CC functional equivalents (e.g. the apo-dystrophin-4 cDNA sequence appearing  
CC as ABK86497). Also included are polynucleotides that hybridise to either  
CC strand of (I), a vector containing (I), a cell containing (I) or the  
CC vector, proteins and peptides encoded by (I), a protein homologous with  
CC human dystrophin that is expressed on cell surfaces in vivo antibodies  
CC (Ab) specific for the protein and method of screening for leukemia cells  
CC by analysing DNA for presence of (I) or by detecting presence of (II).  
CC The apo-dystrophin-4 inversion sequence is a regulatory element that  
CC controls expression (transcription and translation) of associated DNA,  
CC and may allow read-through of stop codons. The apo-dystrophin-4 inversion  
CC sequence is used in gene therapy of diseases associated with truncation  
CC of proteins, particularly muscular dystrophy and also leukaemia, but more  
CC generally (I) is a regulatory sequence used to control expression of any  
CC attached gene. Analysis of DNA for (I), or detection of proteins (II)  
CC encoded by (I), can be used to screen for leukaemic cells and related  
CC diseases. Antibodies raised against (II) can be used therapeutically, to  
CC inhibit (II) activity, also to detect (II) in screening assays. The  
CC present sequence is a apo-dystrophin-4 cDNA fragment showing an inversion  
CC breakpoint (recombination signal sequence) similar to that in with human  
CC dystrophin  
XX  
SQ Sequence 158 BP; 35 A; 35 C; 20 G; 68 T; 0 U; 0 Other;  
Query Match 100.0%; Score 137; DB 6; Length 158;  
Best Local Similarity 100.0%; Pred. No. 4.8e-59;  
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATTATAAGGAAAGAAATAACGCAATGCAAGTGGTGAAGCTGTGAAGTCTCAGGTGT 60  
DB 137 ATTATAAGGAAAGAAATAACGCAATGCAAGTGGTGAAGCTGTGAAGTCTCAGGTGT 78  
QY 61 GCACAAATTATCAGGAACACCCCAAAACCAAAAGTGAAGTGAATATGATGAGAGCCGTG 120  
DB 77 GCACAAATTATCAGGAACACCCCAAAACCAAAAGTGAAGTGAATATGATGAGAGCCGTG 18  
QY 121 TTTGATGTTAATTAATT 137  
DB 17 TTTGATGTTAATTAATT 1  
RESULT 3  
ABK86468/c  
ID ABK86468 standard; DNA; 200 BP.  
XX AC ABK86468;  
XX 27-AUG-2002 (first entry)  
XX Human dystrophin genomic DNA with inversion breakpoint #2.  
XX Human; ds; apo-dystrophin-4; inversion sequence; gene therapy;  
XX protein truncation; muscular dystrophy; leukaemia; dystrophin.  
XX Homo sapiens.  
XX Key Location/Qualifiers  
XX misc\_recomb 42  
XX



```
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATTATTAAGGAAAGAAATACCAATGGCAAGTGTGTAAGCTGTGAAGTCAAGGTGT 60
Db 155 ATTATTAAGGAAAGAAATACCAATGGCAAGTGTGTAAGCTGTGAAGTCAAGGTGT 96
QY 61 GCACAATTATCAGGAACACCCCAACCAACCAAGTGTAGTAATAGCATGAGAACCCGTG 120
Db 95 GCACAATTATCAGGAACACCCCAACCAACCAAGTGTAGTAATAGCATGAGAACCCGTG 36
QY 121 TTTCATGTTTAATTAATT 137
Db 35 TTTCATGTTTAATTAATT 19
RESULT 5
ABK86497
ID ABK86497 standard; cDNA; 996 BP.
AC ABK86497;
XX 27-AUG-2002 (first entry)
XX Human Apo-dystrophin-4 cDNA.
XX Human; ss; gene; apo-dystrophin-4; inversion sequence; gene therapy;
XX protein truncation; muscular dystrophy; leukaemia.
XX Homo sapiens.
PH Key Location/Qualifiers
FT 1..996
FT /tag= a
FT /product= "Apo-dystrophin-4 protein appearing as
FT AAU98739"
FT /partial
FT /note= "No start or stop codon shown"
FT /transl_except= (pos:7..9,aa:Xaa)
FT /transl_except= (pos:49..51,aa:Xaa)
FT /transl_except= (pos:187..189,aa:Xaa)
FT /transl_except= (pos:202..204,aa:Xaa)
FT /transl_except= (pos:217..219,aa:Xaa)
FT /transl_except= (pos:253..255,aa:Xaa)
FT /transl_except= (pos:265..270,aa:Xaa-Xaa)
FT /transl_except= (pos:394..396,aa:Xaa)
FT /transl_except= (pos:547..549,aa:Xaa)
FT /transl_except= (pos:565..567,aa:Xaa)
FT /transl_except= (pos:616..618,aa:Xaa)
FT /transl_except= (pos:649..651,aa:Xaa)
FT /transl_except= (pos:682..684,aa:Xaa)
FT /transl_except= (pos:709..711,aa:Xaa)
FT /transl_except= (pos:790..792,aa:Xaa)
FT /transl_except= (pos:841..843,aa:Xaa)
FT /transl_except= (pos:850..852,aa:Xaa)
FT /transl_except= (pos:880..882,aa:Xaa)
FT /transl_except= (pos:907..909,aa:Xaa)
FT /transl_except= (pos:952..954,aa:Xaa)
FT /transl_except= (pos:982..984,aa:Xaa)
FT /transl_except= (pos:988..990,aa:Xaa)
FT /note= "Xaa= unknown, encoded by in frame stop codon"
XX
PN GB2368064-A.
XX
XX 24-APR-2002.
XX
XX 16-JAN-2001; 2001GB-00001124.
XX
XX 30-SEP-2000; 2000US-0237079P.
XX
XX (IMCR ) IMPERIAL CANCER RES TECHNOLOGY LTD.
XX (BARB/) BARBER E.
XX
XX Barber E;
```

```
XX WPI; 2002-429042/46.
DR P-PSDB; AAU98739.
XX
XX New human regulatory polynucleotide, useful for treating disorders
XX associated with protein truncation, particularly muscular dystrophy, and
XX related peptides and antibodies.
XX
XX Claim 5; Page 170-172; 222pp; English.
XX
XX The invention relates to a polynucleotide (I) comprising, or consisting
XX of, apo-dystrophin-4 inversion sequence appearing as ABK86496, or its
XX functional equivalents (e.g. the apo-dystrophin-4 cDNA sequence appearing
XX as ABK86497). Also included are polynucleotides that hybridise to either
XX strand of (I), a vector containing (I), a cell containing (I) or the
XX vector, proteins and peptides encoded by (I), a protein homologous with
XX human dystrophin that is expressed on cell surfaces in vivo antibodies
XX (Ab) specific for the protein and method of screening for leukemia cells
XX by analysing DNA for presence of (I) or by detecting presence of (II).
XX The apo-dystrophin-4 inversion sequence is a regulatory element that
XX controls expression (transcription and translation) of associated DNA,
XX and may allow read-through of stop codons. The apo-dystrophin-4 inversion
XX sequence is used in gene therapy of diseases associated with truncation
XX of proteins, particularly muscular dystrophy and also leukaemia, but more
XX generally (I) is a regulatory sequence used to control expression of any
XX attached gene. Analysis of DNA for (I) or detection of proteins (II)
XX encoded by (I), can be used to screen for leukaemic cells and related
XX diseases. Antibodies raised against (II) can be used therapeutically, to
XX inhibit (II) activity, also to detect (II) in screening assays. The
XX present sequence is the cDNA sequence for human apo-dystrophin-4
XX containing a plurality of stop codons some of which may be read through
XX due to the presence of (I) in the apo-dystrophin-4 gene
XX
XX Sequence 996 BP; 334 A; 152 C; 195 G; 315 T; 0 U; 0 Other;
SQ
Query Match 100.0%; Score 137; DB 6; Length 996;
Best Local Similarity 100.0%; Pred. No. 4.7e-59;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATTATTAAGGAAAGAAATACCAATGGCAAGTGTGTAAGCTGTGAAGTCAAGGTGT 60
Db 860 ATTATTAAGGAAAGAAATACCAATGGCAAGTGTGTAAGCTGTGAAGTCAAGGTGT 919
QY 61 GCACAATTATCAGGAACACCCCAACCAACCAAGTGTAGTAATAGCATGAGAACCCGTG 120
Db 920 GCACAATTATCAGGAACACCCCAACCAACCAAGTGTAGTAATAGCATGAGAACCCGTG 979
QY 121 TTTCATGTTTAATTAATT 137
Db 980 TTTCATGTTTAATTAATT 996
RESULT 6
ABK86462
ID ABK86462 standard; cDNA; 1230 BP.
XX
XX AC ABK86462;
XX 27-AUG-2002 (first entry)
XX
XX Human Apo-dystrophin-4 full length cDNA.
XX
XX Human; ss; gene; apo-dystrophin-4; inversion sequence; gene therapy;
XX protein truncation; muscular dystrophy; leukaemia.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 1..1230
XX /*tag= a
XX /product= "Apo-dystrophin-4"
XX /partial
XX /note= "No start or stop codon shown"
XX
```



XX SQ Sequence 1234 BP; 405 A; 190 C; 260 G; 379 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 137; DB 6; Length 1234;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-59;  
 Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATAAGGAAAGAAATTAACCAATGGACAAAGTGGTGAAGCTGTGAAGCTGTGAAGCTGTGAAGCTGT 60  
 DB 1098 ATTATAAGGAAAGAAATTAACCAATGGACAAAGTGGTGAAGCTGTGAAGCTGTGAAGCTGT 1157

QY 61 GCACAATTATCAGGAACACCCCAAAACCAAGTGGTGAAGCTGTGAAGCTGTGAAGCTGTGAAGCTGT 120  
 DB 1158 GCACAATTATCAGGAACACCCCAAAACCAAGTGGTGAAGCTGTGAAGCTGTGAAGCTGTGAAGCTGT 1217

QY 121 TTTGATGTTAATAATT 137  
 DB 1218 TTTGATGTTAATAATT 1234

RESULT 8  
 ABK81996/c  
 ID ABK81996 standard; DNA; 2691 BP.  
 AC ABK81996;  
 XX  
 XX  
 XX 13-AUG-2002 (first entry)  
 DT Human dystrophin 3' untranslated region.  
 DE  
 XX  
 XX Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;  
 KW Duchenne's muscular dystrophy; DMD; dystrophin; human; ds.  
 KW  
 XX Homo sapiens.  
 OS  
 XX WO200229056-A2.  
 PN  
 XX 11-APR-2002.  
 PD  
 XX 04-OCT-2001; 2001WO-US031126.  
 PF  
 XX 06-OCT-2000; 2000US-0238848P.  
 PR  
 XX (UNMI ) UNIV MICHIGAN.  
 PA  
 XX Chamberlain JS, Harper SQ;  
 PI  
 XX WPI; 2002-435334/46.  
 DR  
 XX  
 XX A composition for preparing therapeutic drugs, has a mini-dystrophin  
 PT peptide comprising a specific number of spectrin-like repeat domains, or  
 PT a nucleic acid sequence encoding the mini-dystrophin peptide.  
 XX  
 XX Disclosure; Fig 10; 145pp; English.  
 PS  
 XX The invention describes a composition comprising a mini-dystrophin  
 CC peptide comprising a spectrin-like repeat domain, where the domain  
 CC comprises n spectrin-like repeats, and contains no more than n spectrin-  
 CC like repeats, where n is an even number between 4-24, or a nucleic acid  
 CC encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the  
 CC polynucleotide encoding it is useful as a medicament, for preparing a  
 CC drug for therapeutic application and in the preparation of a composition  
 CC for treatment of muscle disease, e.g. Duchenne's muscular dystrophy  
 CC (DMD). This sequence represents a human dystrophin polynucleotide  
 CC sequence used in the creation of the mini-dystrophin peptides of the  
 CC invention  
 XX  
 XX SQ Sequence 2691 BP; 860 A; 448 C; 461 G; 922 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 137; DB 6; Length 2691;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-59;  
 Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATAAGGAAAGAAATTAACCAATGGACAAAGTGGTGAAGCTGTGAAGCTGTGAAGCTGT 60  
 DB 2033 ATTATAAGGAAAGAAATTAACCAATGGACAAAGTGGTGAAGCTGTGAAGCTGTGAAGCTGT 1974

QY 61 GCACAATTATCAGGAACACCCCAAAACCAAGTGGTGAAGCTGTGAAGCTGTGAAGCTGT 120  
 DB 1973 GCACAATTATCAGGAACACCCCAAAACCAAGTGGTGAAGCTGTGAAGCTGTGAAGCTGT 1914

QY 121 TTTGATGTTAATAATT 137  
 DB 1913 TTTGATGTTAATAATT 1897

RESULT 9  
 ABK82000/c  
 ID ABK82000 standard; DNA; 8689 BP.  
 XX  
 XX AC ABK82000;  
 XX  
 XX 13-AUG-2002 (first entry)  
 DT DNA encoding mini-dystrophin protein deltaH2-R19.  
 DE  
 XX  
 XX Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;  
 KW Duchenne's muscular dystrophy; DMD; dystrophin; ds.  
 KW  
 XX Homo sapiens.  
 OS  
 XX Synthetic.  
 OS  
 XX WO200229056-A2.  
 PN  
 XX 11-APR-2002.  
 PD  
 XX 04-OCT-2001; 2001WO-US031126.  
 PF  
 XX 06-OCT-2000; 2000US-0238848P.  
 PR  
 XX (UNMI ) UNIV MICHIGAN.  
 PA  
 XX Chamberlain JS, Harper SQ;  
 PI  
 XX WPI; 2002-435334/46.  
 DR  
 XX  
 XX A composition for preparing therapeutic drugs, has a mini-dystrophin  
 PT peptide comprising a specific number of spectrin-like repeat domains, or  
 PT a nucleic acid sequence encoding the mini-dystrophin peptide.  
 XX  
 XX Disclosure; Fig 15; 145pp; English.  
 PS  
 XX The invention describes a composition comprising a mini-dystrophin  
 CC peptide comprising a spectrin-like repeat domain, where the domain  
 CC comprises n spectrin-like repeats, and contains no more than n spectrin-  
 CC like repeats, where n is an even number between 4-24, or a nucleic acid  
 CC encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the  
 CC polynucleotide encoding it is useful as a medicament, for preparing a  
 CC drug for therapeutic application and in the preparation of a composition  
 CC for treatment of muscle disease, e.g. Duchenne's muscular dystrophy  
 CC (DMD). This sequence represents a mini-dystrophin sequence of the  
 CC invention  
 XX  
 XX SQ Sequence 8689 BP; 2721 A; 1804 C; 1861 G; 2303 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 137; DB 6; Length 8689;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-59;  
 Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATAAGGAAAGAAATTAACCAATGGACAAAGTGGTGAAGCTGTGAAGCTGTGAAGCTGT 60  
 DB 8031 ATTATAAGGAAAGAAATTAACCAATGGACAAAGTGGTGAAGCTGTGAAGCTGTGAAGCTGT 7972

QY 61 GCACAATTATCAGGAACACCCCAAAACCAAGTGGTGAAGCTGTGAAGCTGTGAAGCTGTGAAGCTGT 120  
 DB 7971 GCACAATTATCAGGAACACCCCAAAACCAAGTGGTGAAGCTGTGAAGCTGTGAAGCTGTGAAGCTGT 7912

QY 121 TTGTGATTTAAATTAATT 137  
|||||  
Db 7911 TTGTGATTTAAATTAATT 7895

## RESULT 10

ABK82002/c  
ID ABK82002 standard; DNA; 11443 BP.

AC ABK82002;

XX 13-AUG-2002 (first entry)

XX DNA encoding mini-dystrophin protein deltaR9-R16.

XX Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;

KW Duchenne's muscular dystrophy; DMD; dystrophin; ds.

XX Homo sapiens.

OS Synthetic.

XX WO200229056-A2.

XX 11-APR-2002.

XX 04-OCT-2001; 2001WO-US031126.

XX 06-OCT-2000; 2000US-0238848P.

XX (UNMI ) UNIV MICHIGAN.

XX Chamberlain JS, Harper SQ;

XX WPI; 2002-435334/46.

XX A composition for preparing therapeutic drugs, has a mini-dystrophin  
PT peptide comprising a specific number of spectrin-like repeat domains, or  
PT a nucleic acid sequence encoding the mini-dystrophin peptide.

XX Disclosure; Fig 17; 145pp; English.

XX The invention describes a composition comprising a mini-dystrophin  
CC peptide comprising a spectrin-like repeat domain, where the domain  
CC comprises n spectrin-like repeats, and contains no more than n spectrin-  
CC like repeats, where n is an even number between 4-24, or a nucleic acid  
CC encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the  
CC polynucleotide encoding it is useful as a medicament, for preparing a  
CC drug for therapeutic application and in the preparation of a composition  
CC for treatment of muscle disease, e.g. Duchenne's muscular dystrophy  
CC (DMD). This sequence represents a mini-dystrophin sequence of the  
CC invention

XX Sequence 11443 BP; 3707 A; 2339 C; 2502 G; 2895 T; 0 U; 0 Other;

Query Match 100.0%; Score 137; DB 6; Length 11443;

Best Local Similarity 100.0%; Pred. No. 4.6e-59;

Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATAAGGAAAGAAATAACGCAATGGCAAGTGGTGAAGCTGTAAGTCTCAGGTGT 60

Db 10785 ATTATAAGGAAAGAAATAACGCAATGGCAAGTGGTGAAGCTGTAAGTCTCAGGTGT 10726

QY 61 GCACAATTATCAGGAACACCCCAAAACCAAAAGTAGGTTAGAAATAGCATGAGAAGCCGTG 120

Db 10725 GCACAATTATCAGGAACACCCCAAAACCAAAAGTAGGTTAGAAATAGCATGAGAAGCCGTG 10666

QY 121 TTGTGATTTAAATTAATT 137

Db 10665 TTGTGATTTAAATTAATT 10649

## RESULT 11

## ABK81959/c

ID ABK81959 standard; DNA; 13957 BP.

XX AC ABK81959;

XX 13-AUG-2002 (first entry)

XX cDNA encoding human dystrophin.

XX Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;

KW Duchenne's muscular dystrophy; DMD; dystrophin; human; gene; ds.

XX Homo sapiens.

XX WO200229056-A2.

XX 11-APR-2002.

XX 04-OCT-2001; 2001WO-US031126.

XX 06-OCT-2000; 2000US-0238848P.

XX (UNMI ) UNIV MICHIGAN.

XX Chamberlain JS, Harper SQ;

XX WPI; 2002-435334/46.

XX A composition for preparing therapeutic drugs, has a mini-dystrophin  
PT peptide comprising a specific number of spectrin-like repeat domains, or  
PT a nucleic acid sequence encoding the mini-dystrophin peptide.

XX Example 2; Fig 1; 145pp; English.

XX The invention describes a composition comprising a mini-dystrophin  
CC peptide comprising a spectrin-like repeat domain, where the domain  
CC comprises n spectrin-like repeats, and contains no more than n spectrin-  
CC like repeats, where n is an even number between 4-24, or a nucleic acid  
CC encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the  
CC polynucleotide encoding it is useful as a medicament, for preparing a  
CC drug for therapeutic application and in the preparation of a composition  
CC for treatment of muscle disease, e.g. Duchenne's muscular dystrophy  
CC (DMD). This sequence represents a human dystrophin polynucleotide  
CC sequence used in the creation of the mini-dystrophin peptides of the  
CC invention

XX Sequence 13957 BP; 4602 A; 2781 C; 3122 G; 3452 T; 0 U; 0 Other;

Query Match 100.0%; Score 137; DB 6; Length 13957;

Best Local Similarity 100.0%; Pred. No. 4.6e-59;

Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATAAGGAAAGAAATAACGCAATGGCAAGTGGTGAAGCTGTAAGTCTCAGGTGT 60

Db 13299 ATTATAAGGAAAGAAATAACGCAATGGCAAGTGGTGAAGCTGTAAGTCTCAGGTGT 13240

QY 61 GCACAATTATCAGGAACACCCCAAAACCAAAAGTAGGTTAGAAATAGCATGAGAAGCCGTG 120

Db 13239 GCACAATTATCAGGAACACCCCAAAACCAAAAGTAGGTTAGAAATAGCATGAGAAGCCGTG 13180

QY 121 TTGTGATTTAAATTAATT 137

Db 13179 TTGTGATTTAAATTAATT 13163

## RESULT 12

ABT10904/c

ID ABT10904 standard; cDNA; 13957 BP.

XX AC ABT10904;

XX 04-DEC-2002 (first entry)

## RESULT 11





```

Db      13179 TTTGATGTTAATTAATT 13163

RESULT 15
ABS70403/c
ID      ABS70403 standard; cDNA; 13977 BP.
XX
XX      AC      BS70403;
XX
XX      27-NOV-2002 (first entry)
XX
XX      Human bone remodelling gene #60.
XX
XX      Bone remodelling; osteoporosis; human; gene; ss.
XX
XX      Homo sapiens.
XX
XX      US6426186-B1.
XX
XX      30-JUL-2002.
XX
XX      18-JAN-2000; 2000US-00484970.
XX
XX      18-JAN-2000; 2000US-00484970.
XX
XX      (INCY-) INCYTE GENOMICS INC.
XX
XX      Jones KA, Volkmut W, Walker MG;
XX
XX      WPI; 2002-673014/72.
XX
XX      A combination of polynucleotides which are co-expressed with genes known
XX      to be involved in bone remodeling and osteoporosis are useful in an array
XX      for the diagnosis of bone remodeling and osteoporosis associated
XX      disorders.
XX
XX      Claim 1; Col 185-198; 206pp; English.
XX
XX      The invention relates to a combination comprising a number of
XX      substantially purified and isolated polynucleotides which are co-
XX      expressed with genes known to be involved in bone remodelling and
XX      osteoporosis. The invention is used to diagnose disorders associated with
XX      bone remodelling or osteoporosis. ABS70344-ABS70512 represent human bone
XX      remodelling genes of the invention
XX
XX      Sequence 13977 BP; 4596 A; 2765 C; 3120 G; 3453 T; 0 U; 43 Other;

Query Match      85.4%; Score 117; DB 6; Length 13977;
Best Local Similarity 100.0%; Pred. No. 5.3e-49;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATTATAAGAAAAGAAAATAACGCAATGGACAAAGTGGTGAAGCTGTGAAGTCAAGTGT 60
DB      13306 ATTATAAGAAAAGAAAATAACGCAATGGACAAAGTGGTGAAGCTGTGAAGTCAAGTGT 13247

QY      61 GCACAAATTATCAGGAACACCCCAAAACCAAGTAGAGTGAAGTGAAGTCAAGTGAAGCC 117
DB      13246 GCACAAATTATCAGGAACACCCCAAAACCAAGTAGAGTGAAGTGAAGTGAAGTGAAGCC 13190

Search completed: April 6, 2004, 15:10:51
Job time : 66.9329 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 14:12:12 ; Search time 505.195 Seconds  
(without alignments)  
8098.097 Million cell updates/sec

Title: US-09-966-264d-1

Perfect score: 137

Sequence: 1 attataaagaaaagaaaa.....gtgttgatgtaattaatt 137

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 0

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_pbg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	137	100.0	373	10	AW580404
C 2	137	100.0	396	14	H89576
C 3	137	100.0	455	9	AA427831
C 4	137	100.0	462	14	CB045405

C 5	137	100.0	473	14	CA389721
C 6	137	100.0	489	28	AQ015185
C 7	137	100.0	647	10	AW646484
C 8	137	100.0	696	10	AW950480
C 9	137	100.0	710	9	AV725574
C 10	137	100.0	727	12	BG567176
C 11	137	100.0	740	14	CD356811
C 12	137	100.0	745	14	CD357556
C 13	137	100.0	1121	12	BM546012
C 14	136	99.3	743	14	CD110642
C 15	120	87.6	352	10	AW385154
C 16	112	81.8	862	14	CB962272
C 17	111	81.0	353	10	AW580423
C 18	88	64.2	911	12	BI752714
C 19	86	62.8	375	10	AW607064
C 20	57	41.6	551	9	AA621932
C 21	54	39.4	270	14	F25920
C 22	54	39.4	841	10	BF791019
C 23	48	35.0	265	14	F18701
C 24	46	33.6	631	14	CB465228
C 25	43	31.4	641	9	AI528613
C 26	42	30.7	268	10	BB178429
C 27	42	30.7	291	10	BB174122
C 28	42	30.7	291	10	BB339107
C 29	42	30.7	296	10	BB041920
C 30	42	30.7	329	10	BB085328
C 31	42	30.7	390	9	AI324317
C 32	42	30.7	607	10	BE370292
C 33	42	30.7	610	9	AA146038
C 34	42	30.7	643	14	CB850037
C 35	42	30.7	735	14	CB570707
C 36	42	30.7	4437	11	AK036936
C 37	40	29.2	734	10	BF674672
C 38	40	29.2	924	13	BF455062
C 39	32	23.4	210	10	BB171938
C 40	32	23.4	418	13	BY377466
C 41	31	22.6	293	10	BB307434
C 42	30	21.9	651	13	BU338697
C 43	30	21.9	733	13	BU327193
C 44	27	19.7	945	10	BF180441
C 45	25	18.2	306	10	BB086855

#### ALIGNMENTS

RESULT 1  
AW580404/c  
LOCUS AW580404 373 bp mRNA linear EST 16-MAR-2000  
DEFINITION PM2-HT0451-080100-003-h09 HT0451 Homo sapiens cDNA, mRNA sequence.  
ACCESSION AW580404  
VERSION AW580404.1 GI:7255453  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 373)  
AUTHORS HGCP http://www.ludwig.org.br/ORESTES.  
TITLE The FAPESP/LICR Human Cancer Genome Project  
JOURNAL Unpublished (1999)  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM2&ta=PM2-HT0451-080100-003-h09&t3=2000-01-08&t4=1)

CA389721 cs101f03.  
AQ015185 CIT-HSP-2  
AW646484 hi84e10.x  
AW950480 EST162550  
AV725574 AV725574  
BG567176 602589640  
CD356811 AGENCOURT  
CD357556 AGENCOURT  
BM546012 AGENCOURT  
CD110642 AGENCOURT  
AW385154 PM2-HT045  
CB962272 AGENCOURT  
AW580423 PM2-HT045  
BI752714 603028327  
AW607064 PM2-HT045  
AA621932 ng20b10.s  
F25920 HSPD13162 H  
BF791019 602251072  
F18701 HSPD03293 H  
CB465228 726497 NA  
AI528613 mq87c07.x  
BB178429 BB178429  
BB174122 BB174122  
BB339107 BB339107  
BB041920 BB041920  
BB085328 BB085328  
AI324317 mq87c07.y  
BE370292 601222549  
AA146038 mq87c07.x  
CB850037 NRA-1794  
CB570707 AGENCOURT  
AK036936 Mus muscu  
BF674672 602136802  
BX455062 BX455062  
BB171938 BB171938  
BY377466 BY377466  
BB307434 BB307434  
BU338697 603515030  
BU327193 601744569  
BF180441 601805231  
BB086855 BB086855

Seq primer: puc 18 forward  
 High quality sequence stop: 373.  
 Location/Qualifiers  
 1. .373  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /dev\_stage="Adult"  
 /clone\_lib="HT0451"  
 /note="Organ: head, neck; Vector: puc18; Site: 1: SmaI;  
 Site 2: SmaI; A mini-library was made by cloning products  
 derived from ORESTES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the puc 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."

## FEATURES

source

## ORIGIN

Query Match 100.0%; Score 137; DB 10; Length 373;  
 Best Local Similarity 100.0%; Pred. No. 3e-57;  
 Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATTATTAAGGAAAGAAATACCGAATGGACAACTGGTGAAGCTGTGAACCTCAGGTGT 60  
 Db 369 ATTATTAAGGAAAGAAATACCGAATGGACAACTGGTGAAGCTGTGAACCTCAGGTGT 310  
 QY 61 GCACATTATCAGGAACACCCCAACCAACCAAGTAGAGTAGAATAGCATGAGAACCCGTG 120  
 Db 309 GCACATTATCAGGAACACCCCAACCAACCAAGTAGAGTAGAATAGCATGAGAACCCGTG 250  
 QY 121 TTGATGTTTAATTAATT 137  
 Db 249 TTGATGTTTAATTAATT 233

## RESULT 2

H89576 396 bp mRNA linear EST 28-NOV-1995  
 Yw28c12.r1 Morton Fetal Cochlea Homo sapiens cDNA clone  
 IMAGE:253558 5' similar to gb:M18533 DYSTROPHIN (HUMAN);, mRNA  
 sequence.  
 H89576  
 H89576.1 GI:1079922

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

## AUTHORS

1 (bases 1 to 396)  
 Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,  
 Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,  
 Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N., Le, N.,  
 Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
 Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
 Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.,  
 and Marra, M.

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## COMMENT

Generation and analysis of 280,000 human expressed sequence tags  
 Genome Res. 6 (9), 807-828 (1996)  
 9704478  
 889549  
 Contact: Wilson RK  
 Washington University School of Medicine  
 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 High quality sequence stops: 300  
 Source: IMAGE Consortium, LLNL  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Insert Length: 1088 Std Error: 0.00  
 Seq primer: M13RP1  
 High quality sequence stop: 300.

## FEATURES

source

## Location/Qualifiers

1. .396  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="GDB:3891011"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:253558"  
 /tissue type="cochlea"  
 /dev stage="16-22 week fetus"  
 /lab\_host="SOLR cells (kanamycin resistant)"  
 /clone\_lib="Morton Fetal Cochlea"  
 Note: Organ: ear; Vector: Bluescript SK-; Site 1: EcoRI;  
 Site 2: XhoI; Reference: Genomics 23, 42-50 (1994) Cloned  
 unidirectionally. Primer: Oligo dt. Fetal cochlea, normal.  
 3% of inserts <0.5 kb, 56% 0.5-1.0 kb, 7% >1 kb, Uni-ZAP  
 XR Vector. Library constructed by N. Robertson, C. Morton.  
 -5' adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor  
 sequence: 5' CTCGAGTTTCTTTTCTTTT 3'

## ORIGIN

Query Match 100.0%; Score 137; DB 14; Length 396;  
 Best Local Similarity 100.0%; Pred. No. 3e-57;  
 Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATTATTAAGGAAAGAAATACCGAATGGACAACTGGTGAAGCTGTGAACCTCAGGTGT 60  
 Db 264 ATTATTAAGGAAAGAAATACCGAATGGACAACTGGTGAAGCTGTGAACCTCAGGTGT 205  
 QY 61 GCACATTATCAGGAACACCCCAACCAACCAAGTAGAGTAGAATAGCATGAGAACCCGTG 120  
 Db 204 GCACATTATCAGGAACACCCCAACCAACCAAGTAGAGTAGAATAGCATGAGAACCCGTG 145  
 QY 121 TTGATGTTTAATTAATT 137  
 Db 144 TTGATGTTTAATTAATT 128

## RESULT 3

AA427831/c  
 LOCUS  
 DEFINITION  
 IMAGE:773399 5' similar to gb:M18533 DYSTROPHIN (HUMAN);, mRNA  
 sequence.  
 AA427831  
 AA427831.1 GI:2111628

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## PUBMED

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

```

/dev stage="8-9 weeks"
/lab host="DH10B"
/clone lib="Soares total fetus NB2HF8 9w"
polylinker: Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from pooled 8-9 week
(total) fetus material with a Not I - oligo(dT) primer [5'
TGTACCAATCTAGTGGAGCGCGCCCTTAATTTTCTTTTCTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo. "
ORIGIN

Query Match      100.0%; Score 137; DB 9; Length 455;
Best Local Similarity 100.0%; Pred. No. 3e-57;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATTAAGGAAAGAAATTAACCAATGGACAAAGTGTGAAGCTGTGAAGTCAAGTGTGT 60
DB 155 ATTATTAAGGAAAGAAATTAACCAATGGACAAAGTGTGAAGCTGTGAAGTCAAGTGTGT 96
QY 61 GCACAATTATCAGGACACCCCAAAACCAAGTGTAGTAATAGCATGAGAACCCGCTG 120
DB 95 GCACAATTATCAGGACACCCCAAAACCAAGTGTAGTAATAGCATGAGAACCCGCTG 36
QY 121 TTGTGATGTTAATTAATT 137
DB 35 TTGTGATGTTAATTAATT 19

RESULT 4
CB045405/c
LOCUS
DEFINITION
NISC_gc10c01.y1 NCI_CGAP_Col7 Homo sapiens cDNA clone IMAGE:3218281
5', mRNA sequence.
ACCESSION
CB045405
VERSION
CB045405.1 GI:27783692
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 462)
AUTHORS
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
NATIONAL Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
TITLE
Unpublished (1997)
JOURNAL
Contact: Robert Strausberg, Ph.D.
COMMENT
Email: cgapbs@mail.nih.gov
cDNA Library Preparation:
DNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC)
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
info@image.llnl.gov
Plate: LLAM7870 row: E column: 2
Seq primer: M13Rpl reverse primer (ABI).
Location/Qualifiers
1. .462
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3218281"
/tissue_type="juvenile granulosa tumor"
/lab host="DH10B"
/clone lib="NCI CGAP Col7"
/note="Organ: colon; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies."
ORIGIN

```

```

Query Match      100.0%; Score 137; DB 14; Length 462;
Best Local Similarity 100.0%; Pred. No. 3e-57;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATTAAGGAAAGAAATTAACCAATGGACAAAGTGTGAAGCTGTGAAGTCAAGTGTGT 60
DB 432 ATTATTAAGGAAAGAAATTAACCAATGGACAAAGTGTGAAGCTGTGAAGTCAAGTGTGT 373
QY 61 GCACAATTATCAGGACACCCCAAAACCAAGTGTAGTAATAGCATGAGAACCCGCTG 120
DB 372 GCACAATTATCAGGACACCCCAAAACCAAGTGTAGTAATAGCATGAGAACCCGCTG 313
QY 121 TTGTGATGTTAATTAATT 137
DB 312 TTGTGATGTTAATTAATT 296

RESULT 5
CA389721/c
LOCUS
DEFINITION
cs101f03.y3 Human Retinal pigment epithelium/choroid cDNA
(Un-normalized, unamplified): cs Homo sapiens cDNA clone cs101f03
5', mRNA sequence.
ACCESSION
CA389721
VERSION
CA389721.1 GI:24720152
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 473)
AUTHORS
Wistow,G., Bernstein,S.L., Wyatt,M.K., Farris,R.N., Behal,A.,
Touchman,J.W., Bouffard,G., Smith,D. and Peterson,K.
Expressed sequence tag analysis of human RPE/choroid for the
NSIBank Project: Over 6000 non-redundant transcripts, novel genes
and splice variants
Mol. Vis. 8 (4), 205-220 (2002)
JOURNAL
MEDLINE
PUBMED
COMMENT
Contact: Wistow G
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: graeme@helix.nih.gov
Plate: 101 row: f column: 03
Seq primer: M13Rpl reverse primer (ABI).
Location/Qualifiers
1. .473
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="cs101f03"
/tissue_type="RPE/choroid"
/dev stage="Adult"
/lab host="EMDH10B"
/clone lib="Human Retinal pigment epithelium/choroid cDNA
(Un-normalized, unamplified): cs
/note="Organ: Eye; Vector: pCMVSPORT6; Two different donor
eyes (75-80 years old) yielded approximately 600 mg of
dissected RPE/choroid tissue. This in turn yielded 340 ug
of total RNA and 7 ug of mRNA. A directionally cloned cDNA
library in the pCMVSPORT6 vector was constructed at Life
Technologies (Rockville, MD; now part of Invitrogen Corp),
essentially following the protocols of the SuperScript
Plasmid System (Invitrogen Corp).
<http://www.invitrogen.com/>. The library code
designation was cs. For this library, cDNA inserts were
cloned into the NotI/MluI sites of the vector. EST
analysis was performed on the unamplified library at the
NIH Intramural Sequencing Center (NISC)."
FEATURES
source

```

## ORIGIN

Query Match 100.0%; Score 137; DB 14; Length 473;  
 Best Local Similarity 100.0%; Pred. No. 3e-57;  
 Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATAAGGAAAAAGAAAATAACCAATGGCAAGTGGTGAAGCTGTGAAGTGTGAAGTGTGAAGTGTGAAGTGT 60  
 DB 231 ATTATAAGGAAAAAGAAAATAACCAATGGCAATGGCAAGTGGTGAAGTGTGAAGTGTGAAGTGT 172  
 QY 61 GCACAATTATCAGGAACACCCCAACCAACCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGT 120  
 DB 171 GCACAATTATCAGGAACACCCCAACCAACCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGT 112  
 QY 121 TTGTGATGTTTAATTAATT 137  
 DB 111 TTGTGATGTTTAATTAATT 95

## RESULT 6

AQ015185/c  
 LOCUS 489 bp DNA linear GSS 09-JUN-1998  
 DEFINITION CIT-HSP-2310F4.TR CIT-HSP Homo sapiens genomic clone 2310F4,  
 genomic survey sequence.

ACCESSION AQ015185  
 VERSION AQ015185.1 GI:3193921  
 KEYWORDS GSS.

SOURCE Homo sapiens (human)

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 489)  
 Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,  
 Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,  
 Simon,M. and Venter,J.C.

TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map

JOURNAL Building (1998)

COMMENT Other\_GSSs: CIT-HSP-2310F4.TF

Contact: Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: mdadams@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC  
 end search page:  
[http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html).

Seq primer: M13 Reverse  
 Class: BAC ends.

## FEATURES

source  
 1..489  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /clone="2310F4"  
 /sex="Male"  
 /cell\_type="Sperm"  
 /clone\_lib="CIT-HSP"  
 /notes="Vector: pBelobAC11; Site\_1: HindIII; Site\_2:  
 HindIII"

## ORIGIN

Query Match 100.0%; Score 137; DB 28; Length 489;  
 Best Local Similarity 100.0%; Pred. No. 3e-57;  
 Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATAAGGAAAAAGAAAATAACCAATGGCAAGTGGTGAAGTGTGAAGTGTGAAGTGTGAAGTGT 60  
 DB 358 ATTATAAGGAAAAAGAAAATAACCAATGGCAAGTGGTGAAGTGTGAAGTGTGAAGTGT 299  
 QY 61 GCACAATTATCAGGAACACCCCAACCAACCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGT 120

DB 298 GCACAATTATCAGGAACACCCCAACCAACCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGT 239  
 QY 121 TTGTGATGTTTAATTAATT 137  
 DB 238 TTGTGATGTTTAATTAATT 222

## RESULT 7

AW664684/c

LOCUS 647 bp mRNA linear EST 06-APR-2000  
 DEFINITION h184e10.x1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens CDNA clone  
 IMAGE:2979018 3' similar to gb:M18533 DYSTROPHIN (HUMAN);, mRNA  
 sequence.

ACCESSION AW664684

VERSION AW664684.1 GI:7457227

KEYWORDS EST.

SOURCE Homo sapiens (human)

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 647)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

UNPUBLISHED (1997)

CONTACT: Robert Strausberg, Ph.D.

Email: cgapsb-remail.nih.gov

This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.

Seq primer: -40UP from Gibco

High quality sequence stop: 365.

## FEATURES

source  
 1..647  
 /Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2979018"  
 /lab\_host="DH10B"  
 /clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
 /notes="Organ: pooled; Vector: p773D-Pac (Pharmacia) with  
 a modified polylinker; Site\_1: Not 1; Site\_2: Eco RI;  
 Equal amounts of plasmid DNA from three normalized  
 libraries (fetal lung NbHL19W, testis NHT, and B-cell  
 NCI CGAP GCBI) were mixed, and ss circles were made in  
 vitro. Following HAP purification, this DNA was used as  
 tracer in a subtractive hybridization reaction. The driver  
 was PCR-amplified cDNAs from pools of 5,000 clones made  
 from the same 3 libraries. The pools consisted of  
 1.M.A.G.E. clones 297480-302087, 682632-687239,  
 726408-728711, and 729096-731399. Subtraction by Bento  
 Soares and M. Fatima Bonaldo."

## ORIGIN

Query Match 100.0%; Score 137; DB 10; Length 647;  
 Best Local Similarity 100.0%; Pred. No. 3e-57;  
 Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATAAGGAAAAAGAAAATAACCAATGGCAAGTGGTGAAGTGTGAAGTGTGAAGTGTGAAGTGT 60  
 DB 151 ATTATAAGGAAAAAGAAAATAACCAATGGCAAGTGGTGAAGTGTGAAGTGTGAAGTGT 92  
 QY 61 GCACAATTATCAGGAACACCCCAACCAACCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGT 120  
 DB 91 GCACAATTATCAGGAACACCCCAACCAACCAAGTGAAGTGAAGTGAAGTGAAGTGT 32  
 QY 121 TTGTGATGTTTAATTAATT 137  
 DB 31 TTGTGATGTTTAATTAATT 15

## RESULT 8

AW950480/c

```

LOCUS       AW950480                696 bp    mRNA    linear    EST 01-JUN-2000
DEFINITION  EST362550 MAGE resequences, MAGE Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW950480
VERSION     AW950480.1   GI:8140134
SOURCE      EST.
            Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 696)
AUTHORS     Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C.,
            Holt, I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and
            Quakenbush, J.
TITLE       Assessment of gene expression patterns in a model of colon tumor
            metastasis using a 19,200 element cDNA microarray
JOURNAL     Unpublished (2000)
COMMENT     Contact: John Quakenbush
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: johnq@tigr.org
            Plate: 15
            Seq primer: Reverse.
FEATURES             Location/Qualifiers
     source          1..696
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone_lib="MAGE resequences, MAGE"
                     /note="Vector: pBluescriptSKm"
ORIGIN
Query Match      100.0%; Score 137; DB 10; Length 696;
Best Local Similarity 100.0%; Pred. No. 3e-57;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATAAGGAAAAAGAAAATACCAATGCAAGTGGTAGAAGTGTGAAGCTGTAAGTCAAGTGT 60
Db 572 ATTATAAGGAAAAAGAAAATACCAATGCAAGTGGTAGAAGTGTGAAGCTGTAAGTCAAGTGT 513

QY 61 GCACAATTATCAGGAACACCCCAAAACCAAGTGGTAGAAGTGTGAAGTGTGAAGTGTG 120
Db 512 GCACAATTATCAGGAACACCCCAAAACCAAGTGGTAGAAGTGTGAAGTGTGAAGTGTG 453

QY 121 TTTGATGTTAATAATT 137
Db 452 TTTGATGTTAATAATT 436

RESULT 9
AV725574/c
LOCUS       AV725574 HTC Homo sapiens cDNA clone HTCBD10 5', mRNA sequence.
DEFINITION  AV725574
ACCESSION   AV725574
VERSION     AV725574.1   GI:10831099
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 710)
AUTHORS     Gu, Y., Peng, Y., Song, H., Huang, Q., Yang, Y., Gao, G., Xiao, H., Xu, X.,
            Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L.,
            Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R.,
            Chen, J., Chen, Z. and Han, Z.
            Homo sapiens cDNA HIC Clones
            Unpublished (2000)
            Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)

```

```

Fax: 86-21-50801922
Email: hanq@chgc.sh.cn
This clone is available at CHGC in Shanghai.
Location/Qualifiers
1..710
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HTCBD10"
/tissue_type="Hypothalamus"
/dev_stage="Adult"
/lab_host="SOLR"
/clone_lib="HTC"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
ORIGIN
Query Match      100.0%; Score 137; DB 9; Length 710;
Best Local Similarity 100.0%; Pred. No. 3e-57;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATAAGGAAAAAGAAAATACCAATGCAAGTGGTAGAAGTGTGAAGCTGTAAGTCAAGTGT 60
Db 396 ATTATAAGGAAAAAGAAAATACCAATGCAAGTGGTAGAAGTGTGAAGCTGTAAGTCAAGTGT 337

QY 61 GCACAATTATCAGGAACACCCCAAAACCAAGTGGTAGAAGTGTGAAGTGTGAAGTGTG 120
Db 336 GCACAATTATCAGGAACACCCCAAAACCAAGTGGTAGAAGTGTGAAGTGTGAAGTGTG 277

QY 121 TTTGATGTTAATAATT 137
Db 276 TTTGATGTTAATAATT 260

RESULT 10
BG567176/c
LOCUS       BG567176                727 bp    mRNA    linear    EST 10-APR-2001
DEFINITION  BG567176 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4723833 5',
            mRNA sequence.
ACCESSION   BG567176
VERSION     BG567176.1   GI:13574829
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 727)
AUTHORS     NIH-MGC http://mgi.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgabbs@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LICM1583 row: p column: 10
            High quality sequence stop: 694.
            Location/Qualifiers
1..727
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4723833"
/lab_host="NIH MGC 76"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site 1:
SfiI (ggccctcgcc); Site 2: SfiI (ggccattggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCCATTATGCCC-3' and 3' adaptor sequence:

```





Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATTAAGGAAAAGAAAATAACCAATGACCAAGTGGTGAAGCTGTGAAGTCAAGTGT 60  
 DB 260 ATTATTAAGGAAAAGAAAATAACCAATGACCAAGTGGTGAAGCTGTGAAGTCAAGTGT 201

QY 61 GCACAATTATCAGGAACACCCCAACCAACCAAGTGGTGAAGTCAAGTGTGAAGTCAAGTGT 120  
 DB 200 GCACAATTATCAGGAACACCCCAACCAACCAAGTGGTGAAGTCAAGTGTGAAGTCAAGTGT 141

QY 121 TTGTGATGTTAATTAATT 137  
 DB 140 TTGTGATGTTAATTAATT 124

RESULT 13  
 BM546012 1121 bp mRNA linear EST 20-FEB-2002  
 LOCUS AGENCOURT\_6505286 NIH\_MGC\_125 Homo sapiens cDNA clone IMAGE:5588419  
 DEFINITION 5', mRNA sequence.  
 ACCESSION BM546012  
 VERSION BM546012.1 GI:18778623  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 1121)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: Invitrogen  
 cDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
 Plate: LMAW12359 row: h column: 20  
 High quality sequence start: 50  
 High quality sequence stop: 764.  
 Location/Qualifiers  
 1..1121  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5588419"  
 /lab\_host="DH10B"  
 /clone\_lib="NIH MGC 125"  
 /note="Organ: ovary (pool of 3); Vector: pCMV-SPORT6; Site 1: EcoRV (destroyed); Site 2: NotI; RNA source pool of three ovaries, from females ranging in age from 38 to 49 yo. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 2.1 kb, insert size range 1-3.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 036."

ORIGIN  
 Query Match 100.0%; Score 137; DB 12; Length 1121;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-57;  
 Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATTAAGGAAAAGAAAATAACCAATGACCAAGTGGTGAAGCTGTGAAGTCAAGTGT 60  
 DB 514 ATTATTAAGGAAAAGAAAATAACCAATGACCAAGTGGTGAAGCTGTGAAGTCAAGTGT 455

QY 61 GCACAATTATCAGGAACACCCCAACCAACCAAGTGGTGAAGTCAAGTGTGAAGTCAAGTGT 120  
 DB 454 GCACAATTATCAGGAACACCCCAACCAACCAAGTGGTGAAGTCAAGTGTGAAGTCAAGTGT 395

QY 121 TTGTGATGTTAATTAATT 137  
 DB 394 TTGTGATGTTAATTAATT 378

RESULT 14  
 CD110642 743 bp mRNA linear EST 15-MAY-2003  
 LOCUS AGENCOURT\_13995397 NIH\_MGC\_187 Homo sapiens cDNA clone  
 DEFINITION IMAGE:30373580 5', mRNA sequence.  
 ACCESSION CD110642  
 VERSION CD110642.1 GI:30754851  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 743)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits  
 CDNA Library Preparation: CLONTECH Laboratories, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
 Plate: NDCM167 row: j column: 21  
 High quality sequence stop: 578.  
 Location/Qualifiers  
 1..743  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:30373580"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /clone\_lib="NIH MGC 187"  
 /note="Organ: Blood vessels - aorta, basilar and artery; Vector: pDNR-LIB; Site 1: SfiI (ggccattatggcc); Site 2: SfiI (ggcgcctggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCGCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGGCGCGCCGACATG-DT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.4 kb (range 0.5-4.0 kb). 14/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC Library."

ORIGIN  
 Query Match 99.3%; Score 136; DB 14; Length 743;  
 Best Local Similarity 100.0%; Pred. No. 9.5e-57;  
 Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATTAAGGAAAAGAAAATAACCAATGACCAAGTGGTGAAGCTGTGAAGTCAAGTGT 60  
 DB 138 ATTATTAAGGAAAAGAAAATAACCAATGACCAAGTGGTGAAGCTGTGAAGTCAAGTGT 79

QY 61 GCACAATTATCAGGAACACCCCAACCAACCAAGTGGTGAAGTCAAGTGTGAAGTCAAGTGT 120  
 DB 78 GCACAATTATCAGGAACACCCCAACCAACCAAGTGGTGAAGTCAAGTGTGAAGTCAAGTGT 19

QY 121 TTGTGATGTTAATTAATT 136  
 DB 18 TTGTGATGTTAATTAATT 3

RESULT 15  
 AW385154

LOCUS AW385154 352 bp mRNA linear EST 04-FEB-2000  
 DEFINITION PM2-HT0451-281299-001-a07 HT0451 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION AW385154  
 VERSION AW385154.1 GI:6889813  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 352)  
 HGCP <http://www.ludwig.org.br/ORESTES>.  
 TITLE The FAPESP/LICR Human Cancer Genome Project  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 TEL: +55-11-2704922  
 FAX: +55-11-2707001  
 Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM2&st2=PM2-HT0451-281299-001-a07&t3=1999-12-28&t4=1>)  
 Seq primer: puc 18 forward  
 High quality sequence stop: 352.

## FEATURES

FEATURES	SOURCE
<p>1. 352</p> <p>location/coordinates</p> <p>/organism="Homo sapiens"</p> <p>/mol_type="mRNA"</p> <p>/db_xref="taxon:9606"</p> <p>/dev_stage="Adult"</p> <p>/clone_lib="HT0451"</p> <p>/note="Organ: head neck; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORBSTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."</p>	

## ORIGIN

	Query March	87.6%	Score 120;	DB 10;	Length 352;
	Best Local Similarity	100.0%;	Prod. No. 7.9e-49;		
	Matches 120;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	18	AAATAACGCAATGACAAAGTCGGTGAAGCTGTGTGAACCTCACTCAGGTGTGCACAAATTATCAGGAAC	77		
Db	1	AAATAACGCAATGACAAAGTCGGTGAAGCTGTGTGAACCTCAGGTGTGCACAAATTATCAGGAAC	60		
Qy	78	ACCCCAAAACCAAAGTCAGGTAGAAATAGCATGAGAACCCGCTGTTTGTATGTTAAATTAATT	137		
Db	61	ACCCCAAAACCAAAGTCAGGTAGAAATAGCATGAGAACCCGCTGTTTGTATGTTAAATTAATT	120		

Search completed: April 6, 2004, 17:40:07  
Job time : 506.195 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 13:02:17 ; Search time 574.481 Seconds  
(without alignments)  
10336.278 Million cell updates/sec

Title: US-09-966-264d-1

Perfect score: 137  
Sequence: 1 attataaggaagaaagaaaa.....gtgtttgatgttaataatt 137

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Result No.	Score	Query Match	Length	DB ID	Description
C 1	137	100.0	2148	11	G15848
C 2	137	100.0	2563	9	AK129855
C 3	137	100.0	2691	6	AX538618
C 4	137	100.0	4558	9	BC028720
C 5	137	100.0	8889	6	AX538622
C 6	137	100.0	11443	6	AX538634
C 7	137	100.0	13957	6	AX409637
C 8	137	100.0	13957	6	AX538581
C 9	137	100.0	13957	9	HUMDYS
C 10	137	100.0	98056	9	AC006061
C 11	117	85.4	13977	6	AR220819
C 12	65	47.4	912120	2	AC023414
C 13	46	33.6	13887	4	AF070485
C 14	46	33.6	272578	2	AC108338
C 15	46	33.6	272578	2	AC114184
C 16	42	30.7	13815	6	AX306153
C 17	42	30.7	13815	6	AX538582
C 18	42	30.7	13815	10	MUSDYSA
C 19	42	30.7	19307	6	AR093392
C 20	42	30.7	19307	6	AR142592
C 21	42	30.7	189131	10	AL645477
C 22	30	21.9	13575	5	GGDYS
C 23	22	16.1	208367	2	EX682542
C 24	21	15.3	157446	2	AC118861
C 25	20	14.6	98871	9	AL603824
C 26	20	14.6	94605	9	AF410480S2
C 27	20	14.6	114226	2	AC094710
C 28	20	14.6	126467	9	AF487553
C 29	20	14.6	143410	9	AC087636
C 30	20	14.6	145990	10	AC117837
C 31	20	14.6	157216	2	AC008244
C 32	20	14.6	159557	9	AC127898
C 33	20	14.6	160750	9	HSJ177110
C 34	20	14.6	161601	2	AC021147
C 35	20	14.6	162612	2	AC118722
C 36	20	14.6	165660	2	AC027194
C 37	20	14.6	165789	2	AL591029
C 38	20	14.6	166782	2	AC021516
C 39	20	14.6	169524	2	AC023111
C 40	20	14.6	174398	9	AC009988
C 41	20	14.6	177185	9	AC016632
C 42	20	14.6	180702	2	AC121392
C 43	20	14.6	181390	9	AC074237
C 44	20	14.6	182126	2	AC016639
C 45	20	14.6	187466	9	AC016322

# ALIGNMENTS

RESULT 1  
G15848/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE

human STS CHLC UTR\_01924\_M18533.2148 bp DNA linear STS 19-JAN-1996  
sequence tagged site.  
G15848  
G15848.1 GI:1161737  
STS; STS sequence; primer; sequence tagged site.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 2148)  
Murray, J.L., Sheffield, V., Weber, J.L., Duyk, G. and Buetow, K.H.  
Cooperative Human Linkage Center

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

JOURNAL  
COMMENT  
Unpublished (1995)  
Synonyms: UTR\_01924\_M18533, CHLC.UTR\_01924\_M18533.T36152  
Contact: Dr. Jeffrey C. Murray  
UofI  
The University of Iowa  
Department of Pediatrics, Iowa City, IA 52242, USA  
Tel: (319) 356-3508  
Fax: (319) 356-3347  
Email: jeff-murray@uiowa.edu

Primer A: AACGATTTTGGTGTGTTA  
Primer B: GATATCAGCCCAAGAGATG  
STS size: 189  
PCR Profile:

denature: 30 seconds at 94 degrees C  
annealing: 75 seconds at 55 degrees C  
extension: 15 seconds at 72 degrees C  
PCR cycles: 27  
extension: 6 minutes at 72 degrees C

## Protocol:

Template: 30mg genomic DNA  
Primer: each 1.5 pmole  
dNTPs: each 200 uM  
Taq Polymerase: 0.3 units  
Total Vol: 10 uL

## Buffer:

MgCl2: 1.5mM  
KCl: 50mM  
Tris: 10mM  
pH: 8.3

Prepared with primer pairs derived from M18533.

FEATURES  
source

Location/Qualifiers  
1..2148  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
350..538  
350..369  
complement(519..538)

Query Match 100.0%; Score 137; DB 11; Length 2148;  
Best Local Similarity 100.0%; Pred. No. 5.4e-63;  
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTATAAGGAAAAGAAAATACGCAATCGCAAGTGTGAGCTGTGAACCTCAGGTGT 60  
Db 2033 ATTATAAGGAAAAGAAAATACGCAATCGCAAGTGTGAGCTGTGAACCTCAGGTGT 1974  
Qy 61 GCACAATTATCAGGAACACCCCAAAACCAAAAGTGAGGTAGCAATAGCAGAGCCGTG 120  
Db 1973 GCACAATTATCAGGAACACCCCAAAACCAAAAGTGAGGTAGCAATAGCAGAGCCGTG 1914  
Qy 121 TTTGATGTTAATTAATT 137  
Db 1913 TTTGATGTTAATTAATT 1897

RESULT 2  
AK129855/c AK129855 2563 bp mRNA linear PRI 10-SEP-2003  
LOCUS Homo sapiens cDNA FLJ26345 fis, clone HRT03668.  
DEFINITION  
ACCESSION AK129855  
VERSION AK129855.1 GI:34526478  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1  
Ota,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H., Suzuki,Y.,

Hata.H., Nakagawa,K., Mizuno,S., Morinaga,M., Kawamura,M.,  
Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Nishikawa,T.,  
Sugiyama,A., Kawakami,B., Nagai,K., Isogai,T. and Sugano.S.  
NEDO human cDNA sequencing project

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Unpublished  
2 (bases 1 to 2563)  
Sugano,S. and Suzuki,Y.  
Direct Submission  
Submitted (31-JUL-2003) Sumio Sugano, Institute of Medical Science,  
University of Tokyo, Laboratory of Genome Structure, Human Genome  
Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan  
(E-mail: shirokane@u-tokyo.ac.jp, Tel:81-3-5449-5286,  
Fax:81-3-5449-5416)

## COMMENT

NEDO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
Research Association for Biotechnology (RAB); cDNA library  
construction and 5'-end one pass sequencing: Institute of Medical  
Science, University of Tokyo, Laboratory of Genome Structure, Human  
Genome Center; 3'-end one pass sequencing: RAB; clone selection for  
full insert sequencing: RAB and Helix Research Institute.

FEATURES  
source

Location/Qualifiers  
1..2563  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="HRT03668"  
/tissue\_type="heart"  
/clone\_lib="HRT"  
/note="cloning vector: pME18SFL3"

## ORIGIN

Query Match 100.0%; Score 137; DB 9; Length 2563;  
Best Local Similarity 100.0%; Pred. No. 5.4e-63;  
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ATTATAAGGAAAAGAAAATACGCAATCGCAAGTGTGAGCTGTGAACCTCAGGTGT 60  
Db 1905 ATTATAAGGAAAAGAAAATACGCAATCGCAAGTGTGAGCTGTGAACCTCAGGTGT 1846  
Qy 61 GCACAATTATCAGGAACACCCCAAAACCAAAAGTGAGGTAGCAATAGCAGAGCCGTG 120  
Db 1845 GCACAATTATCAGGAACACCCCAAAACCAAAAGTGAGGTAGCAATAGCAGAGCCGTG 1786  
Qy 121 TTTGATGTTAATTAATT 137  
Db 1785 TTTGATGTTAATTAATT 1769

## RESULT 3

AX538618/c AX538618 2591 bp DNA linear PAT 23-NOV-2002  
LOCUS Sequence 38 from Patent WO0229056.  
DEFINITION  
ACCESSION AX538618  
VERSION AX538618.1 GI:25271161  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

1  
Chamberlain,J.S. and Harper,S.Q.  
Mini-dystrophin nucleic acid and peptide sequences  
Patent: WO 0229056-A 38 11-APR-2002.  
THE REGENTS OF THE UNIVERSITY OF MICHIGAN (US)

FEATURES  
source

Location/Qualifiers  
1..2691  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

## ORIGIN

Query Match 100.0%; Score 137; DB 6; Length 2691;  
Best Local Similarity 100.0%; Pred. No. 5.4e-63;

[illegible]

[illegible]



## RESULT 10

AC006061 Homo sapiens X BAC GSHB-19024 (Genome Systems Human BAC Library) linear PRI 01-MAY-2002  
 LOCUS AC006061 complete sequence.  
 DEFINITION AC006061  
 AC006061  
 AC006061.1 GI:4204246  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 1 (bases 1 to 98056)  
 Muzny, D., Aronson, A.D., Bouck, J., Brundage, E., Bunac, C., Chen, Z.,  
 Di, W., Ding, Y., Dugan, S., Durbin, J., Forcum, J., Garcia, C.,  
 Gorrell, J.H., Gorrell, L.L., Hernandez, J., Jackson, L.,  
 Kondejewski, N., Leal, B., Lichtarge, O., Liu, W., Logan, O., Lu, J.,  
 Martinez, C., Oswal, G., Pampall, L.R., Parish, B.J., Perez, L.,  
 Rasid, N.D., Rives, C., Scherer, S.E., Shen, H., Simon, M., Vo, Q.,  
 Williamson, A., Worley, K.C., Yu, W., Zhou, X., Nelson, D., and  
 Gibbs, R.A.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 98056)  
 Worley, K.C.  
 Direct Submission  
 Submitted (26-NOV-1998) Molecular and Human Genetics, Baylor  
 College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 98056)  
 Worley, K.C.  
 Direct Submission  
 Submitted (30-JAN-1999) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 4 (bases 1 to 98056)  
 Worley, K.C.  
 Direct Submission  
 Submitted (02-FEB-1999) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 5 (bases 1 to 98056)  
 Worley, K.C.  
 Direct Submission  
 Submitted (04-FEB-1999) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 6 (bases 1 to 98056)  
 Worley, K.C.  
 Direct Submission  
 Submitted (28-MAR-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 7 (bases 1 to 98056)  
 Worley, K.C.  
 Direct Submission  
 Submitted (07-MAR-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 8 (bases 1 to 98056)  
 Worley, K.C.  
 Direct Submission  
 Submitted (01-MAY-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jan 30, 1999 this sequence version replaced gi:4176317.  
 INFORMATION: <http://gc.bcm.tmc.edu:8088/home.html> or email  
 gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the  
 entire insert of this clone. Overlapping regions of clones are only  
 sequenced and submitted once, so the sequence for the remainder of  
 the insert may be found in the record for the adjacent clones.

Overlapping clones are noted at the beginning and end of the  
 Features listing.

## ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches  
 of a local database that includes entries from dbSTS, GDB, and  
 local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,  
 unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST  
 (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the  
 EGT and cDNA sequences. Genes demonstrate at least two exons  
 flanked by consensus splice sites that maintained sequence  
 continuity across the splice junctions. Sequences that are not  
 identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum  
 standard of double strand coverage with a minimum of 2 clones and 2  
 reads with no ambiguities or 2 chemistries with a minimum of 2  
 clones and 3 reads with no ambiguities. If the sequence quality for  
 a region does not meet this standard, it will be indicated in the  
 annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality  
 standards - estimated error rate less than 1 per 10,000 bases.

FEATURES	Source	Location/Qualifiers
misc_feature	1. 98056	/organism="Homo sapiens"
		/mol_type="genomic DNA"
		/db_xref="taxon:9606"
		/chromosome="X"
		/clone="GSHB-19024"
	1. 1999	/notes="overlaps bases 151178..153177 of clone AC117405"
		/function="clone overlap"
repeat_region	40..82	/rpt_family="L2"
repeat_region	378..507	/rpt_family="L1MD3"
repeat_region	complement(518..570)	/rpt_family="L1MD3"
repeat_region	571..729	/rpt_family="(GGA)n"
repeat_region	730..987	/rpt_family="L1MC4"
repeat_region	complement(991..1400)	/rpt_family="MSTD"
repeat_region	complement(1649..1779)	/rpt_family="L2"
repeat_region	complement(1783..2212)	/rpt_family="MIR"
repeat_region	complement(3626..3656)	/rpt_family="L2R16A"
repeat_region	complement(4713..5074)	/rpt_family="(GA)n"
repeat_region	5702..7108	/rpt_family="THE1B"
repeat_region	complement(8670..8734)	/rpt_family="L2"
repeat_region	complement(9069..9144)	/rpt_family="MLTIG"
repeat_region	complement(9254..9351)	/rpt_family="MLTIG"
repeat_region	complement(9537..9688)	/rpt_family="MIR"
repeat_region	9797..9942	/rpt_family="WER5A"
repeat_region	complement(10403..10429)	/rpt_family="MIR"
repeat_region	complement(10673..10705)	/rpt_family="AT rich"
repeat_region	complement(11460..11755)	/rpt_family="(CA)n"
repeat_region		/rpt_family="Alu5x"





```

Direct Submission
Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 11, 2000 this sequence version replaced gi:6970579.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIGR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: L6731
Center clone name: 767_B_5

----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 185633 bases at least Q40
Consensus quality: 195115 bases at least Q30
Consensus quality: 201837 bases at least Q20
Insert size: 178000; agarose-fp
Insert size: 209120; sum-of-contigs
Quality coverage: 5.8 in Q20 bases; agarose-fp
Quality coverage: 4.9 in Q20 bases; sum-of-contigs

-----
NOTE: This is a 'working draft' sequence. It currently

```

NOTE: This is a 'working draft' sequence. It currently consists of 31 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of 'N', but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

*	1	1043:	contig	of 1043	bp in length
*		1143:	gap	of 100 bp	
*	1044	1144:	contig	of 1081	bp in length
*	1144	2234:	contig	of 100 bp	
*	2225	2324:	gap	of 100 bp	
*	2325	3345:	contig	of 1021	bp in length
*	3345	3445:	gap	of 100 bp	
*	3446	4502:	contig	of 1057	bp in length
*	4503	4602:	gap	of 100 bp	
*	4603	5756:	contig	of 1154	bp in length
*	5757	5856:	gap	of 100 bp	
*	5857	7097:	contig	of 1241	bp in length
*	7098	7197:	gap	of 100 bp	
*	7198	8236:	contig	of 1039	bp in length
*	8237	8336:	gap	of 100 bp	
*	8337	9353:	contig	of 1017	bp in length
*	9354	9453:	gap	of 100 bp	
*	9454	10705:	contig	of 1252	bp in length
*	10706	10805:	gap	of 100 bp	
*	10806	12332:	contig	of 1527	bp in length
*	12333	12432:	gap	of 100 bp	
*	12433	13877:	contig	of 1445	bp in length
*	13878	13977:	gap	of 100 bp	
*	13978	15201:	contig	of 1224	bp in length
*	15202	15301:	gap	of 100 bp	
*	15302	16368:	contig	of 1667	bp in length
*	16369	17068:	gap	of 100 bp	
*	17069	19060:	contig	of 1992	bp in length
*	19061	19600:	gap	of 100 bp	
*	19601	22639:	contig	of 3479	bp in length
*	22640	22739:	gap	of 100 bp	
*	22740	26376:	contig	of 3637	bp in length
*	26377	26476:	gap	of 100 bp	
*	26477	29275:	contig	of 2799	bp in length
*	29276	29375:	gap	of 100 bp	
*	29376	33655:	contig	of 3890	bp in length
*	33656	33665:	gap	of 100 bp	
*	33666	37012:	contig	of 3647	bp in length
*	37013	37112:	gap	of 100 bp	

FEATURES	Location/Qualifiers
* 37113	39390: contig of 2278 bp in length
* 39391	39430: gap of 100 bp
* 39431	43630: contig of 4200 bp in length
* 43631	43790: gap of 100 bp
* 43791	47621: contig of 3831 bp in length
* 47621	47721: gap of 100 bp
* 47722	53852: contig of 6131 bp in length
* 53853	53952: gap of 100 bp
* 53953	63432: contig of 9470 bp in length
* 63433	63522: gap of 100 bp
* 63523	77351: contig of 13839 bp in length
* 77362	77481: gap of 100 bp
* 77482	92485: contig of 15004 bp in length
* 92486	92565: gap of 100 bp
* 92566	106951: contig of 14386 bp in length
* 106952	107051: gap of 100 bp
* 107052	129678: contig of 22627 bp in length
* 129679	129778: gap of 100 bp
* 129779	151427: contig of 21649 bp in length
* 151428	151527: gap of 100 bp
* 151528	181858: contig of 30311 bp in length
* 181839	181938: gap of 100 bp
* 181939	212120: contig of 30182 bp in length.
FEATURES	Location/Qualifiers
source	1. 212120
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="14"
	/map="14"
	/clone="RP11-767B5"
	/clone_lib="RPC1-11 Human Male BAC"
misc_feature	1. 1043
	/note="assembly_fragment"
misc_feature	1144. 2224
	/note="assembly_fragment"
misc_feature	2325. 3345
	/note="assembly_fragment"
misc_feature	3446. 4502
	/note="assembly_fragment"
misc_feature	4603. 5756
	/note="assembly_fragment"
misc_feature	5857. 7097
	/note="assembly_fragment"
misc_feature	7198. 8236
	/note="assembly_fragment"
misc_feature	8337. 9353
	/note="assembly_fragment"
misc_feature	9454. 10705
	/note="assembly_fragment"
misc_feature	10806. 12332
	/note="assembly_fragment"
misc_feature	12433. 13877
	/note="assembly_fragment"
misc_feature	13978. 15201
	/note="assembly_fragment"
misc_feature	15302. 16968
	/note="assembly_fragment"
misc_feature	17069. 19060
	/note="assembly_fragment"
misc_feature	19161. 22639
	/note="assembly_fragment"
misc_feature	22740. 28376
	/note="assembly_fragment"
misc_feature	26477. 29275
	/note="assembly_fragment"
misc_feature	29376. 33265
	/note="assembly_fragment"
misc_feature	33366. 37012
	/note="assembly_fragment"
misc_feature	37113. 39390
	/note="assembly_fragment"
	clone_end:SP6

vector\_side:right  
39491..43690  
/note="assembly\_fragment"  
43791..47621  
/note="assembly\_fragment"  
47722..53852  
/note="assembly\_fragment"  
clone\_end:17  
vector\_side:left  
53953..53422  
/note="assembly\_fragment"  
63523..77361  
/note="assembly\_fragment"  
77462..92465  
/note="assembly\_fragment"  
92566..106951  
/note="assembly\_fragment"  
107052..129678

Query Match 47.4%; Score 65; DB 2; Length 212120;  
Best Local Similarity 100.0%; Pred. No. 3.7e-24;  
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATTAAGGAAAAAATAACGCAATGACCAAGTGTGAGCTGTGAACTCAGGTGT 60  
|||||  
DB 92401 ATTATTAAGGAAAAAATAACGCAATGACCAAGTGTGAGCTGTGAACTCAGGTGT 92460  
|||||

QY 61 GCACA 65  
|||||  
DB 92461 GCACA 92465

RESULT 13  
AF070485/c 13887 bp mRNA linear MAM 09-DEC-1998  
LOCUS  
DEFINITION Canis familiaris dystrophin mRNA, complete cds.  
ACCESSION AF070485  
VERSION AF070485.1 GI:3982750  
KEYWORDS  
SOURCE Canis familiaris (dog)  
ORGANISM Canis familiaris  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
1 (bases 1 to 13887)  
Carville,K.S., Mann,C.J., Schatzberg,S.J. and Wilton,S.D.  
Direct Submission  
Submitted (04-JUN-1998) ANRI, Pathology, University of Western  
Australia, Verdun Street, Nedlands, WA 6018, Australia  
Location/Qualifiers  
1. 13887  
/organism="Canis familiaris"  
/mol\_type="mRNA"  
/strain="bred Golden retriever"  
/db\_xref="taxon:9615"  
<1..289  
290..11332  
/codon\_start=1  
/product="dystrophin"  
/protein\_id="AAC83646.1"  
/db\_xref="GI:3982751"  
/translation="MLMWEVEDCYEREDVQKFTFKYVNAQSFKQKHENLPSDL  
QDGRLLDLEGTGQKPKERKSTHALNNVNAKLVKKNVNDLVNIGSDIVDG  
NHKTLGLIWNIIHVKVNMKNIMAGLQNTSEKILLSWVRQSTRNYPOVNVNFT  
TSVSGGLANLHSHRPDLFDWNVVCOQSAQRLEHAFNAKYLQGLIEKLLDPEDV  
ATYVDPKSLMYITSLFOVLPOQVSIKAIQVEVMLPRPSQVTRREHFOIHOMHSQ  
QIVSLAQGYERAPSPKPKFSYATQAYVTTSDPTSPPLPSQHLETPDKTSGRS  
LTETANLDSQITALEVLSLWSDALQAGGEISNDVEEVEKQHEHTEGMDLTS  
HQRVGNQVLGSLQITGKLSDETEVEQENLNSKRECVLRVASMEKQNLKVL  
MDLQNLQLEKLDNLTKTEERTKMEKLPDIEDLKRQVQKVLQEDLSQVQVRY  
NSLTHVNVVDESSEGHATALEQKVLGDWANI CRWTEDRWLLQDILLKQWRF  
EEQLFSAMLSKEDAVNKIHTTQKQDSEVLNQLKVLKTDLEKKQTDMDKLSL  
NQDLSALKNTVVAHKEAWLNDNAFQWNLVQLEKSAQISOAVTTTPSLTOTTV  
MEVTVVTRHILVKAQBELPPPPQKQRIIVDSEIKKLDVDITELHSHWIRSE

AVLQSPFAIYRKEGNFSDLKEKVNAIEREAKFKFLQDASRSQAQALVEQMVNEGVN  
ADSIKQASQLNSRWIEFCOLLSENLWLEYQNNIIIFYNQLQQLQOMTWTAAENWLT  
OPTTSEPTAIKSQLKIDEINLSALQIOLERLKIQSIALKKEGQPMFLDADVA  
FTNHQVADVQAREKELQTFDSLPMPRYQETWSTILTITWQQSEKLSIQVTVTE  
YDIMEQRLGELQALQSSLOEQQNLNLTSTIKKMSKAPKLSIDISKYQSEEEISGR  
WKLSSQLVEHQCLQEBQAKRKIQNHITLKKWTTEVDFPLKEWPAQDSEILKR  
OLQKRLLVNDIQTIFQSLNSVNEGAKYKNEAEPEFAGRLTELEBELTNDQWYCRQ  
VYARKEALGGDLKTVSLQDISEMEHWTQABEYLERDFEYKTPDELTQATVEMKR  
AKEAQCKEAKVLLTESVNSVIAQPPAAQKALKEKELDTLTITNYQWLCTRLNGCKT  
LEEVAAQWHELLSYLEKANKWLESEVVKLTNTENISGAEETAEVLDLSLENLMOHSED  
NPQIRILAQTLDGGVMDLENELETNSRWREHEAVPROKLEQSLQSOAQIE  
KSUHLIOESSUSSIDKQALAYADKVDAAQPFQAPQEQSLQSDLSHSHELSBEMKXNQK  
ETAQRVLSQIDVAKQLQDVMKFRFLFPQKPAFEQRLQSKMSMLDEWQKHLPAETKS  
VEQEVQSLNHCNVLYKLSSEVSEVMVITGRQIVQKKTENFEKEDERVTALKL  
HYNELGAKVTRERQKQLEKCLKSRKRNKMNALTETWLAETMELTTRSAVEGMSNLD  
SEVANGKATOEIEKQVHLKSVTEVGEALTVLKKKMLVEDKLSLLNSNLTAVTSR  
AEWLNLLLEYQKMETFDQNDVIITNIIQADLDESKKPKQEDLILKLGKEM  
NDIRPKVSTRDQAANLMANRGDCKRVPKISLNRFAAISHRITKTKASIPUKE  
LEQFNSDIQKLSPLSEAPQOQVNLKEEDFNKMSDNEGTVELQORGNLQORITD  
ERKREIKIQQLQTKHNAKLDORSORRKALEISHQYQYKQDADLDCUDDIEK  
KLASLEPPRDERIKIETDRELQKKEELNAVRRQAEGLSGDSGAAMAVEPQILSKRW  
RETSFQAPRRFLNFQIHTVHEESVVMATDMPELTSYVFSTLITHTVSAQSEV  
BELLNAPDLCAQDPEDLFKEBESLKNIKDSLQOISGRIDIHNKTKAALHSAFPAER  
KLOEALSRLOFOMERVNNMYKQGRFDRSVEKWRPHYDMKILNWLTAEOFLKKT  
QIENWEHAKYKYLKELQDGIQOSQVVRVLTATGEEIIQSSKTNDASLQKLGSL  
NLKQEVCKQALAEERKLEBEQKILSEFQDVNEFVLNLEADANVANIPLERNEQL  
KEKLEQVLLAEELPLRQGLKQNLNETGGTVIVSAPLSEEDQKLENKLTQNLQWIK  
VSRNLPEKQBEIEAHVKDQLQLEQNLHLLMLSPINQLEISLSDWKVVTQLQELRAQPGP  
VAVQAQKQDVEGILSKQHLKPKPATQPAKRLLEDLSMDKVVVTQLQELRAQPGP  
LSLLDRVIKSQRMVMDLEDINEMI KOKATLODLEORRPOLEELITAAQNLKNTSN  
QEARITITRIERIQSQWDEVOEHLQNRLOLTEMKLDQSTQWLEKESQVIGQARA  
KLESKEAPYTVDAIQKKTETQKALDLRQWQINDVANDLAKLLRDYSADDTKY  
HMINTENASWAIHKSLSEREALDETHLQFPDLLEKFLAWLLEASTATTANVQD  
ATHKERLLEDSKVRELMQWQDQIEIAHTDIYHNLSDNGQVRSLSGSDDAALL  
QRRLDNNFKWSLRSKLNIRSHLEASQDKWLHLSQELLVQLQKDELRSQAP  
IGGDFPAQVQNDVHRAFKELTKPEVINSTLETVRIFLTEQPLEGLEKLYOBREL  
PPEERAQNTLLPKAAEVNTQWELNVHSAWOKRIDEALRLOELQATBELDLK  
LRQAEVTKSQWQVGDLLIDSLQDHEKVKALSGEITPLKENYVNDLRAQLTITGI  
QUSFYNTLEDLNRWKLQVLEDAIRQLHEAHRDPGPASQHFSLTSVQGPWHERAI  
SPNKVPIYINHTQTTCDHPKMTLYQSLADLNNVFSAYRTAMKURRLQKALCLDL  
LSLSAADCALDQNLKQNDQPMILOINCLTITYDLEQHNHNVNVCJDMCLNW  
LLNVYDRTGRIRLFSFTGIIISLCAHLEDKRYLFQKQVASTGCDRQGLLH  
DSIOIPQOLGEVAFGSGNIBPSVSCFPANNKPEIEAALFLDMWLEPQSMVWLP  
LHRVAAETAKHQAKNICKECPITIGFRYSKHNFYDIQSCFFFSQRVAKGHMHPY  
MVEYCTTTSGEDVDFAKVKNKFKTRYFAKHPRMGLPVQTVLEGDNMPVTULI  
NFWDSAPASSPOLSHDDTHSRHIEHYASRLKMNENSGYLNDISIPNESIDELHL  
IQHYRSLNQEPLSQPRPAQILISLESEERLELERLADLEGRNRLNQLAEDRLQKQ  
QHEHGLSLPSPPEMPTSPQPRDAELIAEAKLQHKGRLEARNQILEDHKQJUE  
SQLHRLQLLEQPOAEKAVNGTTSVSPSTSLQSDSSQPMLLRVVGSQTSSEMGEDL  
LSPQDTSGLLEVEQNLNHSFPSSRGRNTPGKPMREDTW"

3'UTR repeat\_region 11789..11828 /rpt\_type="candem" /rpt\_unit="candem"  
ORIGIN  
Query Match 33.6%; Score 46; DB 4; Length 13887;  
Best Local Similarity 100.0%; Pred. No. 7.8e-14;  
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 72 AGGAACACCCCAACCAAGTAGGAGTAAATAGCATGAGAACCC 117  
DB 13196 AGGAACACCCCAACCAAGTAGGAGTAAATAGCATGAGAACCC 13151  
RESULT 14  
AC108338/c  
AC108338  
DEFINITION Rattus norvegicus clone CH230-114p1, \*\*\* SEQUENCING IN PROGRESS  
LOCUS \*\*\*, 5 unordered pieces.  
ACCESSION AC108338  
VERSION AC108338.4 GI:23101239

Query Match 33.6%; Score 46; DB 4; Length 13887;  
Best Local Similarity 100.0%; Pred. No. 7.8e-14;  
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 AGGAACACCCCAACCAAGTAGGAGTAAATAGCATGAGAACCC 117  
DB 13196 AGGAACACCCCAACCAAGTAGGAGTAAATAGCATGAGAACCC 13151

RESULT 14  
AC108338/c  
AC108338  
DEFINITION Rattus norvegicus clone CH230-114p1, \*\*\* SEQUENCING IN PROGRESS  
LOCUS \*\*\*, 5 unordered pieces.  
ACCESSION AC108338  
VERSION AC108338.4 GI:23101239

KEYWORDS	HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE	Rattus norvegicus (Norway rat)
ORGANISM	Rattus norvegicus
REFERENCE	1 (bases 1 to 272578)
AUTHORS	Muzny,D,Marie., Metzker,M, Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Angiano,D., Anyalebechi,V., Ayagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregiorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Haves,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensheva,L., Loulsged,H., Lozado,R., Lu,X., Ma,J., Maheshwari,M., Mahindratne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mathiney,S., McLeod,M.P., McNeill,T.Z., Meenen,B., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwankwelenh,O., Okunnu,G., Olarnpunagoon,A., Pal,S., Parks,K., Patelnak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C., Pioppier,F., Poindexter,A., Popovic,D., Primus,E., Fu,L.-L., Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Riley,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,J., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,P., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weis,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
REFERENCE	2 (bases 1 to 272578)
AUTHORS	Worley,K.C.
JOURNAL	Direct Submission
TITLE	Submitted (27-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	3 (bases 1 to 272578)
AUTHORS	Rat Genome Sequencing Consortium.
JOURNAL	Submitted (08-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
TITLE	On Sep 18, 2002 this sequence version replaced gi:21737649.
COMMENT	The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas

Center: Baylor College of Medicine	Genome Center
Web site: http://www.hgsc.bcm.tmc.edu/	Center code: BCM
Contact: hgsc-help@bcm.tmc.edu	Project Information
Center project name: GPMY	Summary Statistics
Center clone name: CH230-114P1	Assembly program: Phrap; version 0.990329
Consensus quality: 218776 bases at least Q40	Consensus quality: 22844 bases at least Q30
Consensus quality: 225163 bases at least Q20	Estimated insert size: 245370; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation	
* NOTE: Estimated insert size may differ from sequence length	
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).	
* consists of 5 contigs. The true order of the pieces	
* is not known and their order in this sequence record is	
* arbitrary. Gaps between the contigs are represented as	
* runs of N, but the exact sizes of the gaps are unknown.	
* This record will be updated with the finished sequence	
* as soon as it is available and the accession number will	
* be preserved.	
* 1 257855: contig of 257855 bp in length	
* 257956 257955: gap of unknown length	
* 261779: contig of 3824 bp in length	
* 261879: gap of unknown length	
* 261880 261114: contig of 1235 bp in length	
* 263214: gap of unknown length	
* 263215 263204: contig of 2990 bp in length	
* 266205 266204: gap of unknown length	
* 266305 272578: contig of 6274 bp in length.	
Location/Qualifiers	
1. 272578	
/organism="Rattus norvegicus"	
/mol_type="Genomic DNA"	
/db_xref="taxon:10116"	
/clone="CH230-114P1"	
1. 1828	
/note="wgs contig"	
257956. 259766	
/note="wgs_contig"	
ORIGIN	
Query Match 33.6%; Score 46; DB 2; Length 272578;	
Best Local Similarity 100.0%; Pred. No. 6.9e-14;	
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY 71 CAGGACACCCCAACCAAGTGGTAGAATGAGTGAAGC 116	
DB 239591 CAGGACACCCCAACCAAGTGGTAGAATGAGTGAAGC 239546	
RESULT 15	
AC114184	
LOCUS	
DEFINITION	Rattus norvegicus clone CH230-230F23, *** SEQUENCING IN PROGRESS
AC114184	
ACCESSION	***, 21 unordered pieces.
VERSION	AC114184.3 GI:23605533

HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 279539)

Murphy D. Marie, Metzker M. Lee, Abramson S., Adams C., Alder J., Allen C., Allen H., Albrooks S., Amin A., Anguiano D., Anyalebechi V., Ayegbi A., Ayodeji M., Baca E., Baden H., Baldwin D., Bandaranaike D., Barber M., Barnstead M., Benahmed F., Biswal K., Blair J., Blankenburg K., Blyth P., Brown M., Bryant N., Buhay C., Burch P., Burrell K., Calderon E., Cardenas V., Carter K., Cavazos I., Ceasar H., Center A., Chacko J., Chavez D., Chen G., Chen R., Chen Z., Chu J., Cleveland C., Cockrell R., Cox C., Coyle M., Cree A., D'Souza L., Davila M.L., Davis C., Davy-Carroll L., De Anda C., Dederich D., Delgado O., Denson S., Deramo C., Ding Y., Dinh H., Divya K., Draper H., Dugan-Rocha S., Dunn A., Durbin K., Duval B., Eaves K., Egan A., Escotto M., Evans C., Evans C.A., Falls T., Fan G., Fernandez S., Finley M., Flagg N., Forbes L., Foster M., Foster P., Fraser C.M., Gabisi A., Ganta R., Garcia A., Garner T., Garza M., Gebregorgis E., Geer K., Gill R., Grady M., Guerra W., Guevara W., Gunaratne P., Haaland W., Hamil C., Hamilton C., Hamilton K., Harvey J., Havlak P., Hawes A., Henderson N., Hernandez J., Hernandez R., Hines S., Hladun S.B., Hodgson A., Hogues M., Hollins B., Howells S., Huiyk S., Hume J., Idlebird D., Jackson A., Jackson L., Jacob L., Jiang H., Johnson B., Johnson R., Jolivet A., Karpathy S., Kelly S., Kelly S., Khan Z., King L., Kovar C., Kowals C., Kraft C.L., Lebow H., Levan J., Lewis L., Li Z., Liu J., Liu J., Liu W., Liu Y., London P., Longacre S., Lopez J., Lohsenuwa L., Loulseged H., Lozano R.J., Lu X., Ma J., Maheshwari M., Mahindartne M., Mahmoud M., Malloy K., Mangun A., Mangum B., Mapua P., Martin K., Martin R., Martinez E., Mathew S., McLeod M.P., McNeill T.Z., Meenen E., Milosavljevic A., Miner G., Minja E., Montemayor J., Moore S., Morgan M., Morris K., Morris S., Munidasa M., Murphy M., Nair L., Nankervis C., Neal D., Newton N., Nguyen N., Norris S., Parks K., Pasoketeme O., Okunolu G., Olarnpunsagoon A., Pal S., Parks K., Paoletti S., Paul H., Perez A., Perez L., Pflanzko C., Plopper F., Poindexter A., Popovic D., Primus E., Pu L.-L., Puazo M., Quirroz J., Rachlin E., Reeves K., Regier M.A., Reigh R., Reilly B., Reilly M., Ren Y., Reuter M., Richards S., Riggs F., Rivas C., Rodkey T., Rojas A., Rose M., Rose R., Ruiz S.J., Sanders W., Saverly G., Scherer S., Scott G., Shatsman S., Shen H., Shetty J., Shwartsbeyn A., Sisson I., Sitter C.D., Smajic D., Snead A., Sodergren E., Song X.-Z., Sorelle R., Sosa J., Steimle M., Strong R., Sutton A., Svatek A., Taber P., Taylor C., Taylor T., Thomas N., Thomas S., Tingey A., Trejos Z., Usmani K., Valas R., Vera V., Villasana D., Waldron L., Walker B., Wang J., Wang Q., Wang S., Warren J., Warren R., Wei X., White F., Williams G., Willson R., Wiczysk R., Wooden H., Worley K., Wright D., Wright R., Wu J., Yakub S., Yen J., Yoon L., Yoon V., Yu F., Zhang J., Zhou J., Zhou X., Zhao S., Dunn D., von Niederhausern A., Weiss R., Smith D.R., Holt R.A., Smith H.O., Weinstock G. and Gibbs R.A.

Direct Submission

Unpublished

2 (bases 1 to 279539)

Worley K.C.

Direct Submission

Submitted (07-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 279539)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (11-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Oct 9, 2002 this sequence version replaced gi:21733949.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas

(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with 'N' to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GLPV  
Center clone name: CH230-230F23  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 212054 bases at least Q40  
Consensus quality: 219191 bases at least Q30  
Consensus quality: 223518 bases at least Q20  
Estimated insert size: 217504; sum-of-contigs estimation  
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
\* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 21 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 6869: contig of 6869 bp in length  
6870 6969: gap of unknown length  
122455: contig of 115486 bp in length  
122456 122555: gap of unknown length  
122556 185205: contig of 62650 bp in length  
185206 185305: gap of unknown length  
185306 189675: contig of 4370 bp in length  
189676 189775: gap of unknown length  
189776 240091: contig of 50316 bp in length  
240092 240191: gap of unknown length  
240192 245854: contig of 5663 bp in length  
245855 245984: gap of unknown length  
245985 247058: contig of 1102 bp in length  
247059 247156: gap of unknown length  
247157 248831: contig of 1675 bp in length  
248832 248931: gap of unknown length  
250272 250372: contig of 1341 bp in length  
250373 252290: gap of unknown length  
252291 252390: contig of 1918 bp in length  
252391 254224: contig of 1834 bp in length  
254225 254324: gap of unknown length  
254325 255684: contig of 1360 bp in length  
255685 255784: gap of unknown length  
255785 258317: contig of 2533 bp in length  
258318 258417: gap of unknown length  
258418 261382: contig of 2965 bp in length  
261383 261482: gap of unknown length  
261483 264268: contig of 2686 bp in length  
264269 264565: gap of unknown length  
264566 267565: contig of 3297 bp in length  
267566 271534: contig of 3869 bp in length  
271535 271634: gap of unknown length  
271635 275008: contig of 3374 bp in length  
275009 275108: gap of unknown length  
275109 277043: contig of 1935 bp in length

Search completed: April 6, 2004, 16:30:22  
Job time : 581.481 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 03:22:40 ; Search time 4175.64 Seconds  
(without alignments)  
10338.454 Million cell updates/sec

Title: US-09-966-264D-2

Perfect score: 996

Sequence: 1 ggggttgattgatagataaa.....gtgttgatgtaataatt 996

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.mu.\*

20: em.om.\*

21: em.or.\*

22: em.ov.\*

23: em.pat.\*

24: em.ph.\*

25: em.pl.\*

26: em.ro.\*

27: em.sts.\*

28: em.un.\*

29: em.vi.\*

30: em.htg.hum.\*

31: em.htg.in.\*

32: em.htg.other.\*

33: em.htg.mus.\*

34: em.htg.pln.\*

35: em.htg.rod.\*

36: em.htg.mam.\*

37: em.htg.vrt.\*

38: em.sy.\*

39: em.htgo.hum.\*

40: em.htgo.mus.\*

41: em.htgo.other.\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	859.8	86.3	98056	9	AC006061	AC006061 Homo sapi
C 2	858.2	86.2	212120	2	AC023414	AC023414 Homo sapi
C 3	788.2	77.1	212120	2	AC023414	AC023414 Homo sapi
C 4	435	43.7	189131	10	AL645477	AL645477 Mouse DNA
5	431.4	43.3	272578	2	AC108338	AC108338 Rattus no
6	412.2	41.4	2463	6	AX817316	AX817316 Sequence
7	412.2	41.4	4658	9	BC028720	BC028720 Homo sapi
8	410.8	41.2	5339	6	AX538620	AX538620 Sequence
9	410.8	41.2	5417	6	AX538619	AX538619 Sequence
10	410.8	41.2	5462	6	AX538621	AX538621 Sequence
11	410.8	41.2	8689	6	AX538622	AX538622 Sequence
12	410.8	41.2	11443	6	AX538624	AX538624 Sequence
13	410.8	41.2	12057	6	AX538627	AX538627 Sequence
14	410.8	41.2	12446	9	HSDMDR	X14298 Human mRNA
15	410.8	41.2	13957	6	AX409637	AX409637 Sequence
16	410.8	41.2	13957	6	AX538581	AX538581 Sequence
17	410.8	41.2	13957	9	HUMDYS	M18533 Homo sapien
18	409.2	41.1	13977	6	AR220819	AR220819 Sequence
C 19	402.4	40.4	279539	2	AC114184	AC114184 Rattus no
20	396.8	39.8	2148	11	G15848	G15848 human STS C
21	396.8	39.8	2691	6	AX538618	AX538618 Sequence
22	374.2	37.6	13815	6	AX306153	AX306153 Sequence
23	374.2	37.6	13815	6	AX538582	AX538582 Sequence
24	374.2	37.6	13815	10	MUSDYSA	M68859 Mouse dystz
C 25	374.2	37.6	19307	6	AR093392	AR093392 Sequence
C 26	374.2	37.6	19307	6	AR142592	AR142592 Sequence
27	368	36.9	13887	4	AF070485	AF070485 Canis fam
28	350	35.1	350	9	AF213444	AF213444 Homo sapi
C 29	319.8	32.1	56285	2	AC100170	AC100170 Mus muscu
30	290.4	29.2	2110	9	HUMDMXX	M92650 Human Duch
31	282.4	28.4	13575	5	GGDYS	X13389 Chicken mRN
32	278.8	28.0	2563	9	AK129855	AK129855 Homo sapi
33	260	26.1	3163	6	E30223	E30223 Shortened d
34	260	26.1	3172	6	E30222	E30222 Shortened d
35	260	26.1	4075	6	E30221	E30221 Shortened d
36	260	26.1	4402	6	E30219	E30219 Shortened d
37	260	26.1	4402	6	E30220	E30220 Shortened d
38	252.8	25.4	3747	6	E30218	E30218 Shortened d
39	218	21.9	218	9	HUMDYS20	M6903 H. sapiens d
40	189.6	19.0	468	5	AF375546	AF375546 Xenopus l
41	188.4	18.9	509	5	AF082741	AF082741 Gallus ga
42	166.2	16.7	1534	5	AF375547	AF375547 Scyliorhi
C 43	151.4	15.2	2148	11	G15848	G15848 human STS C
C 44	151.4	15.2	2563	9	AK129855	AK129855 Homo sapi
C 45	151.4	15.2	2691	6	AX538618	AX538618 Sequence

ALIGNMENTS

RESULT 1

AC006061/c

LOCUS AC006061 98056 bp DNA linear PRI 01-MAY-2002

DEFINITION Homo sapiens X BAC GSHB-19024 (Genome Systems Human BAC Library)

complete sequence.

AC006061

ACCESSION AC006061.1 GI:4204246

VERSION HTG.

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 98056)

AUTHORS Muzny,D., Arenson,A.D., Bouck,J., Brundage,E., Bunac,C., Chen,Z., Di,W., Ding,Y., Dugan,S., Durbin,J., Forcum,J., Garcia,C.,

Pred. No. is the number of results predicted by chance to have a

Gorrell,J.H., Gorrell,L.L., Hernandez,J., Jackson,L.,  
Kondejewski,N., Leal,B., Lichtarge,O., Liu,W., Logan,O., Lu,J.,  
Martinez,C., Oswal,G., Pampell,L.R., Parish,B.J., Perez,L.,  
Rashid,N.D., Rives,C., Scherer,S.E., Shen,H., Simon,M., Vo,Q.,  
Williamson,A., Worley,K.C., Yu,W., Zhou,X., Nelson,D. and  
Gibbs,R.A.

# Direct Submission

Unpublished  
2 (bases 1 to 98056)

## Worley,K.C.

Direct Submission  
Submitted (26-NOV-1998) Molecular and Human Genetics, Baylor  
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 98056)

## Worley,K.C.

Direct Submission  
Submitted (30-JAN-1999) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

4 (bases 1 to 98056)

## Worley,K.C.

Direct Submission  
Submitted (04-FEB-1999) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

6 (bases 1 to 98056)

## Worley,K.C.

Direct Submission  
Submitted (02-FEB-1999) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

5 (bases 1 to 98056)

## Worley,K.C.

Direct Submission  
Submitted (04-FEB-1999) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

7 (bases 1 to 98056)

## Worley,K.C.

Direct Submission  
Submitted (28-MAR-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

7 (bases 1 to 98056)

## Worley,K.C.

Direct Submission  
Submitted (07-MAR-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

8 (bases 1 to 98056)

## Worley,K.C.

Direct Submission  
Submitted (01-MAY-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

On Jan 30, 1999 this sequence version replaced gi:4176317.

INFORMATION: <http://gc.bcm.tmc.edu:8088/home.html> or email

gc-help@bcm.tmc.edu

## CLONE LENGTH:

This sequence does not necessarily represent the  
entire insert of this clone. Overlapping regions of clones are only  
sequenced and submitted once, so the sequence for the remainder of  
the insert may be found in the record for the adjacent clones.

## Overlapping clones

are noted at the beginning and end of the  
Features listing.

## ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches  
of a local database that includes entries from dbSTS, GDB, and  
local mapping efforts

Repeats are identified using RepeatMasker (A. Smit and P. Green,  
unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST  
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the  
EST and cDNA sequences. Genes demonstrate at least two exons  
flanked by consensus splice sites that maintained sequence  
continuity across the splice junctions. Sequences that are not  
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum  
standard of double strand coverage with a minimum of 2 clones and 2  
reads with no ambiguities or 2 chemistries with a minimum of 2  
clones and 3 reads with no ambiguities. If the sequence quality for  
a region does not meet this standard, it will be indicated in the  
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality  
standards - estimated error rate less than 1 per 10,000 bases.

## FEATURES

FEATURES	Location/Qualifiers	Source
misc_feature	1..1299 /notes="overlaps bases 151178..153177 of clone AC117405"	1..98056
repeat_region	40..82 /rpt_family="L2"	1..98056
repeat_region	378..507 /rpt_family="L1MD3"	1..98056
repeat_region	complement(518..570) /rpt_family="(GGA)n"	1..98056
repeat_region	571..729 /rpt_family="L1MC4"	1..98056
repeat_region	730..987 /rpt_family="WSTD"	1..98056
repeat_region	complement(991..1400) /rpt_family="L2"	1..98056
repeat_region	complement(1849..1779) /rpt_family="MIR"	1..98056
repeat_region	complement(1783..2212) /rpt_family="LTR16A"	1..98056
repeat_region	complement(3626..3656) /rpt_family="(GA)n"	1..98056
repeat_region	complement(4713..5074) /rpt_family="THE1B"	1..98056
repeat_region	5702..7108 /rpt_family="L2"	1..98056
repeat_region	complement(8670..8734) /rpt_family="MLTIG"	1..98056
repeat_region	complement(9069..9144) /rpt_family="MLTIG"	1..98056
repeat_region	complement(9254..9351) /rpt_family="MIR"	1..98056
repeat_region	complement(9537..9688) /rpt_family="MER5A"	1..98056
repeat_region	9797..9942 /rpt_family="MIR"	1..98056
repeat_region	complement(10403..10429) /rpt_family="AT-rich"	1..98056
repeat_region	complement(10673..10705) /rpt_family="(CA)n"	1..98056
repeat_region	complement(11460..11755) /rpt_family="AluSx"	1..98056
repeat_region	1197..12125 /rpt_family="MER5A"	1..98056
repeat_region	12556..13643 /rpt_family="L1MD2"	1..98056
repeat_region	13618..13808 /rpt_family="L1MD3"	1..98056
repeat_region	complement(13997..14297) /rpt_family="MER4D"	1..98056
repeat_region	1437..14625 /rpt_family="AluSg"	1..98056
repeat_region	complement(14648..14915) /rpt_family="MER4D"	1..98056
repeat_region	complement(14915..15147) /rpt_family="MER4D"	1..98056
repeat_region	15286..15404 /rpt_family="FLAM_C"	1..98056





## JOURNAL

Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

## COMMENT

On Apr 11, 2000 this sequence version replaced gi:6970579.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L6731

Center clone name: 767 B.5

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 185633 bases at least Q40

Consensus quality: 195115 bases at least Q30

Consensus quality: 201837 bases at least Q20

Insert size: 178000; agarose-fp

Quality coverage: 5.8 in Q20 bases; agarose-fp

Quality coverage: 4.9 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently consists of 31 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1      1043: contig of 1043 bp in length
*      1044: gap of 100 bp
*      1144: contig of 1081 bp in length
*      2224: gap of 100 bp
*      2225: gap of 100 bp
*      3345: contig of 1021 bp in length
*      3346: gap of 100 bp
*      3446: contig of 1057 bp in length
*      4503: gap of 100 bp
*      4603: contig of 1154 bp in length
*      5757: gap of 100 bp
*      5758: gap of 100 bp
*      7097: contig of 1241 bp in length
*      7098: gap of 100 bp
*      7198: contig of 1039 bp in length
*      8237: gap of 100 bp
*      8337: contig of 1017 bp in length
*      9354: gap of 100 bp
*      9454: contig of 1252 bp in length
*      10706: gap of 100 bp
*      10806: contig of 1527 bp in length
*      12333: gap of 100 bp
*      12433: contig of 1445 bp in length
*      13878: gap of 100 bp
*      13978: contig of 1224 bp in length
*      15202: gap of 100 bp
*      15302: contig of 1667 bp in length
*      16968: gap of 100 bp
*      17069: contig of 1992 bp in length
*      19061: gap of 100 bp
*      19161: contig of 3479 bp in length
*      22639: gap of 100 bp
*      22640: contig of 3637 bp in length
*      26377: gap of 100 bp
*      26378: contig of 2799 bp in length
*      29376: gap of 100 bp
*      29377: contig of 3890 bp in length
*      33266: gap of 100 bp
*      33267: contig of 3647 bp in length
*      37013: gap of 100 bp
*      37113: contig of 2278 bp in length

```

## FEATURES

## source

```

1. 212120
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="14"
/map="14"
/clone="RP11-767B5"
/clone_lib="RPC1-11 Human Male BAC"
/notes="assembly_fragment"
1. 1043
1144. 2224
/notes="assembly_fragment"
2325. 3345
/notes="assembly_fragment"
3446. 4502
/notes="assembly_fragment"
4603. 5756
/notes="assembly_fragment"
5857. 7097
/notes="assembly_fragment"
7198. 8236
/notes="assembly_fragment"
8337. 9353
/notes="assembly_fragment"
9454. 10705
/notes="assembly_fragment"
10806. 12332
/notes="assembly_fragment"
12433. 13877
/notes="assembly_fragment"
13978. 15201
/notes="assembly_fragment"
15302. 16968
/notes="assembly_fragment"
17069. 19060
/notes="assembly_fragment"
19161. 22639
/notes="assembly_fragment"
22740. 26376
/notes="assembly_fragment"
26477. 29275
/notes="assembly_fragment"
29376. 33265
/notes="assembly_fragment"
33366. 37012
/notes="assembly_fragment"
37113. 39390
/notes="assembly_fragment"
clone_end:SP6
vector_side:right"

```

```

*      39391
*      39491
*      43691
*      43791
*      43791
*      47721
*      47722
*      53853
*      53853
*      53953
*      63423
*      63523
*      77362
*      77462
*      92466
*      92566
*      106952
*      107052
*      129679
*      129779
*      151428
*      151527
*      181839
*      181939

```

## Location/Qualifiers

```

1. 212120
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="14"
/map="14"
/clone="RP11-767B5"
/clone_lib="RPC1-11 Human Male BAC"
/notes="assembly_fragment"
1. 1043
1144. 2224
/notes="assembly_fragment"
2325. 3345
/notes="assembly_fragment"
3446. 4502
/notes="assembly_fragment"
4603. 5756
/notes="assembly_fragment"
5857. 7097
/notes="assembly_fragment"
7198. 8236
/notes="assembly_fragment"
8337. 9353
/notes="assembly_fragment"
9454. 10705
/notes="assembly_fragment"
10806. 12332
/notes="assembly_fragment"
12433. 13877
/notes="assembly_fragment"
13978. 15201
/notes="assembly_fragment"
15302. 16968
/notes="assembly_fragment"
17069. 19060
/notes="assembly_fragment"
19161. 22639
/notes="assembly_fragment"
22740. 26376
/notes="assembly_fragment"
26477. 29275
/notes="assembly_fragment"
29376. 33265
/notes="assembly_fragment"
33366. 37012
/notes="assembly_fragment"
37113. 39390
/notes="assembly_fragment"
clone_end:SP6
vector_side:right"

```





**TITLE**  
JOURNAL  
**COMMENT**

Direct Submission  
Submitted (29-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk Clone request version: clonerequest@sanger.ac.uk  
On Apr 29, 2002 this sequence version replaced gi:17304537.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e. phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em: ENBL; Sw:  
SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP  
database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep RP23-155L18 is  
from the RPCI-23 Mouse PAC Library  
constructed by the group of Pieter de Jong.  
For further details see http://www.chori.org/bacpac/home.htm  
VECTOR: pBACe3.6

----- Genome Center  
Center: UK Medical Research Council  
Center code: UK-MRC  
Web site: http://mrcseq.har.mrc.ac.uk  
Contact: mouseq@har.mrc.ac.uk  
-----

FEATURES	Location/Qualifiers
source	1..189131 /organism="Mus musculus" /mol_type="genomic DNA" /db_xref="taxon:10090" /chromosome="X" /clone_lib="RP23-155L18" /clone_lib="RPCI-23"

**ORIGIN**

Query Match	43.7%; Score 435; DB 10; Length 189131;
Best Local Similarity	84.7%; Pred. No. 1.8e-78;
Matches 577; Conservative	0; Mismatches 80; Indels 24; Gaps 7;
Qy 192	TTTTTCCATGAACACCAGTAGGCTCATCAACATTAAAGTCATATACTAACAACCAACCT 251
Db 113722	TCTCCTCCATCAATGACACCATGCTGATCCGATTTAAGCTAATACTAAC ----ACCA 113776
Qy 252	GTAATGGCTTCATTAATAAGCGTTTGTTCTTCTGTGAACACTGGTGAAAAATCAAACCTT 311
Db 113777	TGCAATGTGTCATTAAACAAGATTGTGTTCTTGTGAATAATGGTGTAAAA --CGGACTGT 113834
Qy 312	GTTGTGTACCCCTGCATGAGCTTCTGTGTCTTCCACCAAGAATGGGAATGATTT 371
Db 113835	GGTCTGTATACCTTCAATGCAGCTTATGTGTGTCTTTCC --TGAATGGTATGACTC 113892
Qy 372	CCCAATGGCAAGAAACAGAGTGATGCTATCTATCTGACACCTTTTGTAAAAGTCTGCTT 431
Db 113893	CCAATAGTGGCAACCCAGGGGTACAATACT -----TGCACACTTTGTAA ---CTCTT 113941
Qy 432	TCTTCTCTTTGTTTTCCAGACACAATGTAGNAGTCTTTTCCATGTCAGATGATTT 491
Db 113942	TCTTCTCTTTGTTTTCCAGACACAATGTAGNAGCTTTTCCATGTCAGATGATTT 114001
Qy 492	GGGACAGCGATGGAGTCTTTAGTATCAGTCATGACAGATGAAGAGGACAGATAAAT 551
Db 114002	GGGACAGCGATGGAGTCTTTAGTATCAGTCATGACAGATGAAGAGGACAGATAAAT 114061
Qy 552	GTTTACAACTCCTGATTCGCCGATGGTTTTTATATATTTCATACAAAGAGGATTAG 611

RESULT 4	AL645477	189131 bp	DNA	linear	ROD 29-JUN-2002
LOCUS	AL645477				
DEFINITION	Mouse DNA sequence from clone Rp23-155118 on chromosome X, complete sequence.				
ACCESSION	AL645477				
VERSION	AL645477.8	GI:20338482			
KEYWORDS	HTG.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Rodentia; Sciurognath.; Muridae; Murinae; Mus.				
REFERENCE	1 (bases 1 to 189131)				
AUTHORS	Lovell, J.				

```

Db 114062 GTTTTACAACTCCTGATCCCGCATGGTTTTTATATATTCGTACACAAAGAGGATTAG 114121
QY 612 ACAGTAAGAGTTTACAAAGAAAT-AAATCTATATTTTGTGAAGGCTAGTGGTATTATAT 670
Db 114122 ACAGTAAGAGTTTACAAAGAAATTAATCTATATTTTGTGAAGGCTAGTGGTACTAT 114181
QY 671 TGAGATTTCAAGTAGTTTCTAAGTCGTGTTATGTTTGTGTTTGTACAAATGGCAGGTTTACACG 730
Db 114182 GTAGATTTTCAAGTAGTTTCTAAGTCGTGTTATGTTTGTGTTTGTACAAATGGCAGGTTTACACG 114241
QY 731 TCTATGCAATTTGACAAAAAGTTTAAAGAACTACATGTAATCTTGTAGTAGTAAAT 790
Db 114242 TCTATGCAATTTGACAAAAAGTTTAAAGAAA--ACATGTAATCTTGTAGTAGTAAAT 114298
QY 791 AACTTGCCATTTCTTATATGGAAGCAGTATTTGGTGTGTTAAAAATTTATAACAGTTAT 850
Db 114299 AACTTGCCATTTCTTATATGGAAGCAGTATTTGGGTGTTAAAAATTTATAACAGTTAT 114358
QY 851 AAAGAAGAATTTATAAGGAA 871
Db 114359 AAAGAAGAATTTGTAACATAAA 114379

RESULT 5
AC108338 272578 bp DNA linear HTG 08-OCT-2002
LOCUS Rattus norvegicus clone CH230-114P1, *** SEQUENCING IN PROGRESS
DEFINITION ***; 5 unordered pieces.
ACCESSION AC108338
VERSION AC108338.4 GI:23101239
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 272578)
Muzny,D,Marie, Metzker,M,Lee, Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davilla,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Hawlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpach,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensheva,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Maindarte,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,S., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokeleneh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfamknoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,

```

```

Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villaseana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wlezyk,R., Woodson,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 272578)
Worley,K.C.
Direct Submission
Submitted (27-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 272578)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (08-OCT-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 18, 2002 this sequence version replaced gi:21737649.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GPMV
Center clone name: CH230-114P1
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 218776 bases at least Q40
Consensus quality: 22844 bases at least Q30
Consensus quality: 225163 bases at least Q20
Estimated insert size: 245370; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 257855: contig of 257855 bp in length
* 257856 257955: gap of unknown length
* 257956 261775: contig of 3824 bp in length
* 261780 261875: gap of unknown length
* 261880 263114: contig of 1235 bp in length
* 263115 263214: gap of unknown length
* 263215 266204: contig of 2990 bp in length

```

LOCUS AX817316 2463 bp DNA linear PAT 10-DEC-2003  
DEFINITION Sequence 64 from Patent WO02081517.  
ACCESSION AX817316  
VERSION AX817316.1 GI:39722703  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Decristofaro, M.F., Padigaru, M., Miller, C., Tchernev, V., Zhong, H.,  
Zhong, M., Anderson, D., Ballinger, R., Gerlach, V., Spytek, K.A.,  
Rastelli, L., Kekuda, R., Guo, X., Zethusen, B., Andrew, D., Mezes, P.,  
Patturajan, M., Burgess, C.E., Eisen, A., Wolenc, A., Baumgartner, J.,  
Shimkets, R.A., Gusev, V., Vernet, C.A., Taupier, R.J., Pera, C.,  
Shenoy, S., Li, L., Casman, S., Bolgog, F., Fernandes, E., Smithson, G.,  
Malyankar, U., Tailon, B. and Liu, X.  
TITLE Novel polypeptides and nucleic acids encoded thereby  
JOURNAL Patent: WO 02081517-A 64 17-OCT-2002;  
Curagen Corporation (US)  
FEATURES  
source Location/Qualifiers  
1..2463  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
CDS  
1..2108  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAE84363.1"  
/db\_xref="GI:39722704"  
ORIGIN  
41.4%; Score 412.2; DB 6; Length 2463;  
Query Match 97.0%; Pred. No. 1.5e-73;  
Best Local Similarity 0; Mismatches 13; Indels 0; Gaps 0;  
Matches 420; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
QY 439 CTTTGTTCAGGACACCAATGTAGGAAGTCTTTTCCACATGGCAGATGTTTGGCAGA 498  
DB 1998 CCTAGTTCAAGAGGACCAATGTAGGAAGTCTTTTCCACATGGCAGATGTTTGGCAGA 2057  
QY 499 GCGATGGAGTCTTAGTATCAGTCATGACAGATGAAGAGGAGCAGATTAATGTTTAC 558  
DB 2058 GCGATGGAGTCTTAGTATCAGTCATGACAGATGAAGAGGAGCAGATTAATGTTTAC 2117  
QY 559 AACTCCTGATTCGCGCATGTTTATTAATATTCATACAAAGAGGATTAGACAGTAA 618  
DB 2118 AACTCCTGATTCGCGCATGTTTATTAATATTCATACAAAGAGGATTAGACAGTAA 2177  
QY 619 GAGTTTACAGAAATAATCTATATTTTGTGAAGGAGTGGTATTATCTAGTAGATT 678  
DB 2178 GAGTTTACAGAAATAATCTATATTTTGTGAAGGAGTGGTATTATCTAGTAGATT 2237  
QY 679 CAGTAGTTTCTAAGTCTGTTATTGTTTGTAAAGTGGCAGGTTTACAGTCTATGCA 738  
DB 2238 CAGTAGTTTCTAAGTCTGTTATTGTTTGTAAAGTGGCAGGTTTACAGTCTATGCA 2297  
QY 739 ATTGTACAAAAAGTTATAGAAACACATGTAATCTTTAGTAGCTTAATAATACCTGCC 798  
DB 2298 ATTGTACAAAAAGTTATAGAAACACATGTAATCTTTAGTAGCTTAATAATACCTGCC 2357

\* 266205 266304: gap of unknown length  
\* 266305 272578: contig of 6274 bp in length.  
FEATURES  
source Location/Qualifiers  
1..272578  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-114P1"  
misc\_feature 1..1828  
/note="wgs contig"  
misc\_feature 257956..259766  
/note="wgs\_contig"  
ORIGIN  
Query Match 43.3%; Score 431.4; DB 2; Length 272578;  
Best Local Similarity 80.5%; Pred. No. 9e-78;  
Matches 611; Conservative 0; Mismatches 121; Indels 27; Gaps 8;  
QY 117 TGATGAATCTTGCTTATCTATCGAATCTTTTGATATATATTAATGAGAA-CC 175  
DB 237426 TGATGAAGCCTTATACCGCTCTATGAAGTCCCTTTGAGATGTTTAGAAGAGGAATGC 237485  
QY 176 TGAATGTAGC---TTGACATTTTCCATGTAACACAGTAGCTGATCCACATTAAGC 232  
DB 237486 TGAATGCATCAATTATACATTTTCCCATCAATGAGCGCACGCTGATCCAGCATTAAGC 237545  
QY 233 TGATACTAACAAACACAGTGAATGCTTCAATTAATGAAGCTTTGCTTTCTCTGGAAC 292  
DB 237546 TAATACCAAC-----TCCATGCAATGCTTCACTAAACAGGCTTTGCTGCTGTAGAAAT 237600  
QY 293 TGGTGAATAATCAACCTTGCTGTGACACCTCGATGAGCTTCTGCTGCTGCTTCCACC 352  
DB 237601 GGT---AAACTGGATGCTGCTGCTGATACCTTCAATGAGCTTCTGCTGCTTCTTCC 237657  
QY 353 CAGAAATGGGAATGATTTCCCAATGGCAAGAAACAGAGTGATGCTATCTATCTGCAC 412  
DB 237658 ---AGAAATGAATGATTTCCATTTGGCAACCA-----GGGCTACAATGCTGTCAC 472  
QY 413 CTTTGTGAAGTCTGCTTTCTTTCTTTGTTTCCAGACACATGAGGAGTCTTT 532  
DB 237707 ACTTTGTAAGCTCTTTTCTTTCTTTGTTT---CAGACACAATGAGGAGCTTT 237765  
QY 473 TCCACATGGCAGATGATTTGGGAGAGCGATGAGTCTTTAGTATCAGTCATGACAGATG 532  
DB 237766 TCCACATGGCAGATGATTTGGGAGAGCGATGAGTCTTTAGTATCAGTCATGATAGTG 237825  
QY 533 AGAAGAGCAGAAATAATGTTTCAACTCTGATTCGCGATGCTTTTATATATTC 592  
DB 237826 AGAAGAGCAGAAATAATGTTTCAACTCTGATTCGCGATGCTTTTATATATTC 237885  
QY 593 ATACAAAGAGAGGATTAGACAGTGAAGTGTACAAAGAAATAATCTATATTTGTGAA 652  
DB 237886 GTACACAAAGAGGATTAGACAGTGAAGTGTACAAAGAAATAATCTATATTTGTGAA 237945  
QY 653 GGGTAGTGGTATTAATCTAGTATGATTTTCAAGTCTGTTTATGTTTGTAACT 712  
DB 237946 GGGTAGTGGTACTATATCTAGTATGATTTTCAAGTCTGTTTATGTTTGTAACT 238005  
QY 713 AATGGCAGGTTTACACGCTGATGCAATGCTACAAAGAAAGTATTAAGAAACCTACATGA 772  
DB 238006 AATGGCAGGTTTACACGCTGATGCAATGCTACAAAGAAAGTATTAAGAAACCTACATGA 238062  
QY 773 AAATCTGTAGTAAATAACTTGCATTTCTTTATATGAAACGCAATTTGGTGTGTTA 832  
DB 238063 AAATCTGTAGTAAATAACTTGCATTTCTTTATATGAAACGCAATTTGGTGTGTTA 238122  
QY 833 AAATTTATACAGTATTAAGAAAGTATTAAGAA 871  
DB 238123 AAATTTATACAGTATTAAGAAAGTATTAAGAA 238161  
RESULT 6  
AX817316

799 ATTCTTTTATGGAACGATTTTGGTTGGTTTAAAAATTTTAAACAGTTTATAAGAAAG 858  
 |||||  
 Db 2358 ATTCTTTTATGGAACGATTTTGGTTGGTTTAAAAATTTTAAACAGTTTATAAGAAAG 2417  
 |||||  
 QY 859 AATTATAAGGAA 871  
 |||||  
 Db 2418 ATTGTAACATAA 2430  
 |||||

RESULT 7  
 BC028720 4658 bp mRNA linear PRI 25-AUG-2003  
 LOCUS Homo sapiens dystrophin (muscular dystrophy, Duchenne and Becker  
 DEFINITION types), transcript variant Dp71b, mRNA (CDNA clone IMAGE:4822807),  
 complete cds.  
 ACCESSION BC028720  
 VERSION  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, P.,  
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
 Stapleton, T.E., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,  
 Carninci, P., Prange, C., Raha, S.S., Loguercio, N.A., Peters, G.J.,  
 Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,  
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
 Vallalao, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
 Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,  
 Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
 Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
 Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smal, D.E.,  
 Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
 Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED 12477932  
 REFERENCE 2 (bases 1 to 4658)  
 PUBLISHED  
 REFERENCE  
 AUTHORS Strausberg, R.  
 TITLE Submitted (29-APR-2002) National Institutes of Health, Mammalian  
 JOURNAL Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 REMARK NTH-MGC Project URL: <http://mgc.nci.nih.gov>  
 COMMENT Contact: MGC help desk  
 Email: [cgaps@mail.nih.gov](mailto:cgaps@mail.nih.gov)  
 Tissue Procurement: Miklos Falkovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki  
 Toshiyuki and Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbio.org>  
 contact: [amadane@systemsbio.org](mailto:amadane@systemsbio.org)  
 Anup Madan, Jessica Fahey, Erin Helton, Mark Kettman, Anuradha  
 Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAC Plate: 46 Row: 1 Column: 3  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 5032296  
 This clone has the following problem: The cds is short compared to  
 the longest cds in the locus.

FEATURES  
source

Location/Qualifiers  
 1..4658  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone\_image="4822807"  
 /tissue\_type="Testis"  
 /clone\_lib="NIH MGC\_97"  
 /lab\_host="DH10B"  
 /note="Vector: pBluescript"

## gene

1..4658  
 /gene="DMD"  
 /note="synonyms: BMD, DXS142, DXS164, DXS206, DXS230,  
 DXS239, DXS268, DXS269, DXS270, DXS272"  
 /db\_xref="LocusID:1756"  
 /db\_xref="MIM:300377"  
 130..2037  
 /codon\_start=1  
 /product="dystrophin, Dp71b isoform"  
 /protein\_id="AA428720.1"  
 /db\_xref="GI:20379676"  
 /db\_xref="LocusID:1756"

## CDS

130..2037  
 /product="dystrophin, Dp71b isoform"  
 /protein\_id="AA428720.1"  
 /db\_xref="GI:20379676"  
 /db\_xref="LocusID:1756"  
 /translation="MREQLKHETQTTWMDHPKWTLYOSLADLNVRFSAYATMKL  
 RLKQALCLDLISLAAACDALQDNKLNKQNDQMDLIQINCLTKTYRLEQERNLVN  
 VPLQDMCLNLLNVDYDTRGIRVLSPKTIISLKAKHEDKYFLKQVASTGCF  
 CDORGLLLHDSIQIPLQGEVASFSGNSIPEVSCFQFANNKEIIEALFLDWR  
 LEQSRMLVPLHVARAATAKQAKCNICEKPIIGFYSKHKFNVDYICQCFPSG  
 RVAKGHMYPMEYCTPTSSGDDVDFAKLVKNKTKRYFAKHPRMGYLPQVTLK  
 GDMNETVTLINFWFSDPASPFLSHDTHSRISYARSLAEENSGSYLNDIS  
 PNESIDDEHLIIQYCSNLQDPLSQPRSPQILISLESEERGELEILADLEENR  
 NLQAEYDLKQHEHKLSPLEPPPEMPTSPQSPRAELIAKALLRQHKLEARM  
 QILEDHNKQLESQHLRLQLEQPAEKVNGTVSSPSTLQSRSSQPMILRVVGS  
 QTSDEEDLLSPQDTSTGLEVMQLNNSPSSRGHNVGLSFHMADDLGRAMESL  
 VSVMTDEEGAE"

## misc\_feature

844..981  
 /note="ZZ; Region: Zinc finger, ZZ type. Zinc finger  
 present in dystrophin, CBP/p300. ZZ in dystrophin binds  
 calmodulin Putative zinc finger"  
 /db\_xref="CDD:pfam00569"

## ORIGIN

Query Match 41.4%; Score 412.2; DB 9; Length 4658;  
 Best Local Similarity 97.0%; Pred. No. 1.3e-73;  
 Matches 420; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
 QY 439 CTTTGTTCAGGACACAAATGTAGGAAGTCTTTTCACATGGCAGATGTTGGGCGA 498  
 |||||  
 Db 1927 CCTAGTTCAAGAGGACACAAATGTAGGAAGTCTTTTCACATGGCAGATGTTGGGCGA 1986  
 |||||  
 QY 499 CGATGGAGTCTTAGTATCATGTCATGACAGATGAAGAAGGAGAGATAAATGTTTAC 558  
 |||||  
 Db 1987 CGATGGAGTCTTAGTATCATGTCATGACAGATGAAGAAGGAGAGATAAATGTTTAC 2046  
 |||||  
 QY 559 AACTCCTGATTCGGCATGGTTTATAATATTCATAACAAGAGAGATTAGACAGTAA 618  
 |||||  
 Db 2047 AACTCCTGATTCGGCATGGTTTATAATATTCATAACAAGAGAGATTAGACAGTAA 2106  
 |||||  
 QY 619 GAGTTTACAAGAAATAATCTATATTTTGTGAAGGAGTAGTGGTATTACTGTAGATTT 678  
 |||||  
 Db 2107 GAGTTTACAAGAAATAATCTATATTTTGTGAAGGAGTAGTGGTATTACTGTAGATTT 2166  
 |||||  
 QY 679 CAGTAGTTTCTAAGTCTGTTATGTTTGTAAACAATGGCAGGTTTACACGTCCTATGCA 738  
 |||||  
 Db 2167 CAGTAGTTTCTAAGTCTGTTATGTTTGTAAACAATGGCAGGTTTACACGTCCTATGCA 2226  
 |||||  
 QY 739 ATTGTACAAAAAGTTTAAAGAAACATCATGTAAAAATCTTTGATGCTAATAAATCTTGCC 798  
 |||||  
 Db 2227 ATTGTACAAAAAGTTTAAAGAAACATCATGTAAAAATCTTTGATGCTAATAAATCTTGCC 2286  
 |||||  
 QY 799 ATTCTTTTATGGAACGATTTTGGTTGGTTTAAAAATTTTAAACAGTTTATAAGAAAG 858  
 |||||  
 Db 2287 ATTCTTTTATGGAACGATTTTGGTTGGTTTAAAAATTTTAAACAGTTTATAAGAAAG 2346  
 |||||



QY	859	AATTATAAGGAA	871
Db	2347	ATTGTAAACTAAA	2359
RESULT 8			
AX538620	LOCUS	AX538620	5339 bp DNA linear PAT 23-NOV-2002
DEFINITION	Sequence 40 from Patent WO0229056.		
ACCESSION	AX538620		
VERSION	AX538620.1 GI:25271166		
KEYWORDS	synthetic construct		
SOURCE	artificial sequences.		
ORGANISM	Chamberlain, J.S. and Harper, S.Q.		
REFERENCE	Mini-dystrophin nucleic acid and peptide sequences		
AUTHORS	Patent: WO 0229056-A 40 11-APR-2002;		
TITLE	THE REGENTS OF THE UNIVERSITY OF MICHIGAN (US)		
JOURNAL	Location/Qualifiers		
FEATURES	1..5339		
source	/organism="Synthetic"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:32630"		
	/note="Synthetic"		
ORIGIN	Query Match	41.2%; Score 410.8; DB 6; Length 5339;	
	Best Local Similarity	99.3%; Pred.No. 2.5e-73;	
	Matches 415; Conservative	0; Mismatches 7; Indels 0; Gaps 0;	
QY	450	AGGACACAATGTAGGAAGTCTTTTCACATGGCAGATGATTTGGGCAGACCGATGGAGTC	509
Db	4491	AGGACACAATGTAGGAAGTCTTTTCACATGGCAGATGATTTGGGCAGACCGATGGAGTC	4550
QY	510	CTTAGTATCATGTCATGCACAGATGAAGAAGGAGCAGATAAATGTTTTACAACCTCCTGATT	569
Db	4551	CTTAGTATCATGTCATGCACAGATGAAGAAGGAGCAGATAAATGTTTTACAACCTCCTGATT	4610
QY	570	CCGCATGGTTTTTATAATTCATCAACAAGAGAGGATTAGACAGTAAGATTTCACAAAG	629
Db	4611	CCGCATGGTTTTTATAATTCATCAACAAGAGGATTAGACAGTAAGATTTCACAAAG	4670
QY	630	AAATAAATCATATTTTTTGTAAGGGPAGTGGTATTAATCTGTAGATTTTCAGTAGTTTCT	689
Db	4671	AAATAAATCATATTTTTTGTAAGGGPAGTGGTATTAATCTGTAGATTTTCAGTAGTTTCT	4730
QY	690	AAGTCGCTTATGTTTGTAAACAATGGCAGGTTTTACACGCTCATGCAATTCGCACAAA	749
Db	4731	AAGTCGCTTATGTTTGTAAACAATGGCAGGTTTTACACGCTCATGCAATTCGCACAAA	4790
QY	750	AAGTTATAAGAAACTACATGTAATAATCTTTGATAGCTAAATAACTTGCCCAATTTCTTTATA	809
Db	4791	AAGTTATAAGAAACTACATGTAATAATCTTTGATAGCTAAATAACTTGCCCAATTTCTTTATA	4850
QY	810	TGGAACGCATTTTGGGTTGTTTTAAAAATTTATAACAGTTATAAGAAGAATTTATAAAGG	869
Db	4851	TGGAACGCATTTTGGGTTGTTTTAAAAATTTATAACAGTTATAAGAAGAATTTATAAAGG	4910
QY	870	AA 871	
Db	4911	AA 4912	
RESULT 9			
AX538619	LOCUS	AX538619	5417 bp DNA linear PAT 23-NOV-2002
DEFINITION	Sequence 39 from Patent WO0229056.		
ACCESSION	AX538619		
VERSION	AX538619.1 GI:25271163		
KEYWORDS	synthetic construct		
SOURCE			

## ORIGIN

Query Match 41.2%; Score 410.8; DB 6; Length 5462;  
 Best Local Similarity 98.3%; Pred. No. 2.5e-73;  
 Matches 415; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 450 AGGACACAATGTAGGAAGCTTTTCCACATGGCAGATGATTTGGCAGAGCGATGGAGTC 509  
 DB 4614 AGGACACAATGTAGGAAGCTTTTCCACATGGCAGATGATTTGGCAGAGCGATGGAGTC 4673

QY 510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGATAAATGTTTTTCAACTCCTGATT 569  
 DB 4674 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGATAAATGTTTTTCAACTCCTGATT 4733

QY 570 CCGCATGCTTTTATTAATTTATACACAAAGAGGATTAGACAGTAAAGTTTACAAG 629  
 DB 4734 CCGCATGCTTTTATTAATTTATACACAAAGAGGATTAGACAGTAAAGTTTACAAG 4793

QY 630 AAATAAATCTATATTTTGTGAAGGAGTGTGTTATATCTAGATTTTCACTAGTTTCT 689  
 DB 4794 AAATAAATCTATATTTTGTGAAGGAGTGTGTTATATCTAGATTTTCACTAGTTTCT 4853

QY 690 AGTCTGTTATTTGTTTGTAAATTTGAGGAGGTTTACAGTCTATGCAATTTGTACAAA 749  
 DB 4854 AGTCTGTTATTTGTTTGTAAATTTGAGGAGGTTTACAGTCTATGCAATTTGTACAAA 4913

QY 750 AAGTTATAGAAACTACATGTAATAATCTTGATAGCTAAATAACTTGCCTTTTATA 809  
 DB 4914 AAGTTATAGAAACTACATGTAATAATCTTGATAGCTAAATAACTTGCCTTTTATA 4973

QY 810 TGAACGCAATTTGGTGTGTTTAAATTTTAAACAGTTTAAAGAAAGTATTAAGG 869  
 DB 4974 TGAACGCAATTTGGTGTGTTTAAATTTTAAACAGTTTAAAGAAAGTATTAAGG 5033

QY 870 AA 871  
 DB 5034 AA 5035

## RESULT 11

AX538622 LOCUS 8689 bp DNA linear PAT 23-NOV-2002  
 DEFINITION Sequence 42 from Patent WO0229056.  
 ACCESSION AX538622  
 VERSION AX538622.1 GI:25271171  
 KEYWORDS synthetic construct  
 SOURCE synthetic construct  
 ORGANISM artificial sequences.

REFERENCE 1  
 AUTHORS Chamberlain, J.S. and Harper, S.Q.  
 TITLE Mini-dystrophin nucleic acid and peptide sequences  
 JOURNAL Patent: WO 0229056-A 42 11-APR-2002;  
 THE REGENTS OF THE UNIVERSITY OF MICHIGAN (US)

FEATURES  
 source  
 1. .8689  
 /organism="synthetic construct"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:32630"  
 /note="Synthetic"

## ORIGIN

Query Match 41.2%; Score 410.8; DB 6; Length 8689;  
 Best Local Similarity 98.3%; Pred. No. 2.3e-73;  
 Matches 415; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 450 AGGACACAATGTAGGAAGCTTTTCCACATGGCAGATGATTTGGCAGAGCGATGGAGTC 509  
 DB 5985 AGGACACAATGTAGGAAGCTTTTCCACATGGCAGATGATTTGGCAGAGCGATGGAGTC 6044

QY 510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGATAAATGTTTTTCAACTCCTGATT 569  
 DB 6045 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGATAAATGTTTTTCAACTCCTGATT 6104

QY 570 CCGCATGCTTTTATTAATTTATACACAAAGAGGATTAGACAGTAAAGTTTACAAG 629  
 DB 6105 CCGCATGCTTTTATTAATTTATACACAAAGAGGATTAGACAGTAAAGTTTACAAG 6164

QY 630 AAATAAATCTATATTTTGTGAAGGAGTGTGTTATATCTAGATTTTCACTAGTTTCT 689  
 DB 6165 AAATAAATCTATATTTTGTGAAGGAGTGTGTTATATCTAGATTTTCACTAGTTTCT 6224

QY 690 AAGTCTGTTATTTGTTTGTAAATTTGAGGAGGTTTACAGTCTATGCAATTTGTACAAA 749  
 DB 6225 AAGTCTGTTATTTGTTTGTAAATTTGAGGAGGTTTACAGTCTATGCAATTTGTACAAA 6284

QY 750 AAGTTATAGAAACTACATGTAATAATCTTGATAGCTAAATAACTTGCCTTTTATA 809  
 DB 6285 AAGTTATAGAAACTACATGTAATAATCTTGATAGCTAAATAACTTGCCTTTTATA 6344

QY 810 TGAACGCAATTTGGTGTGTTTAAATTTTAAACAGTTTAAAGAAAGTATTAAGG 869  
 DB 6345 TGAACGCAATTTGGTGTGTTTAAATTTTAAACAGTTTAAAGAAAGTATTAAGG 6404

QY 870 AA 871  
 DB 6405 AA 6406

RESULT 12  
 AX538624 LOCUS 11443 bp DNA linear PAT 23-NOV-2002  
 DEFINITION Sequence 44 from Patent WO0229056.  
 ACCESSION AX538624  
 VERSION AX538624.1 GI:25271175  
 KEYWORDS synthetic construct  
 SOURCE synthetic construct  
 ORGANISM artificial sequences.

REFERENCE 1  
 AUTHORS Chamberlain, J.S. and Harper, S.Q.  
 TITLE Mini-dystrophin nucleic acid and peptide sequences  
 JOURNAL Patent: WO 0229056-A 44 11-APR-2002;  
 THE REGENTS OF THE UNIVERSITY OF MICHIGAN (US)

FEATURES  
 source  
 1. .11443  
 /organism="synthetic construct"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:32630"  
 /note="Synthetic"

ORIGIN

Query Match 41.2%; Score 410.8; DB 6; Length 11443;  
 Best Local Similarity 98.3%; Pred. No. 2.2e-73;  
 Matches 415; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 450 AGGACACAATGTAGGAAGCTTTTCCACATGGCAGATGATTTGGCAGAGCGATGGAGTC 509  
 DB 8739 AGGACACAATGTAGGAAGCTTTTCCACATGGCAGATGATTTGGCAGAGCGATGGAGTC 8798

QY 510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGATAAATGTTTTTCAACTCCTGATT 569  
 DB 8799 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGATAAATGTTTTTCAACTCCTGATT 8858

QY 570 CCGCATGCTTTTATTAATTTATACACAAAGAGGATTAGACAGTAAAGTTTACAAG 629  
 DB 8859 CCGCATGCTTTTATTAATTTATACACAAAGAGGATTAGACAGTAAAGTTTACAAG 8918

QY 630 AAATAAATCTATATTTTGTGAAGGAGTGTGTTATATCTAGATTTTCACTAGTTTCT 689  
 DB 8919 AAATAAATCTATATTTTGTGAAGGAGTGTGTTATATCTAGATTTTCACTAGTTTCT 8978

QY 690 AAGTCTGTTATTTGTTTGTAAATTTTAAACAGTTTAAAGAAAGTATTAAGG 749  
 DB 8979 AAGTCTGTTATTTGTTTGTAAATTTTAAACAGTTTAAAGAAAGTATTAAGG 9038

QY	750	AGGTTATAGAAAACTACATGTAATAAATCTTGATAGCTTAATAAATCTGCCATTTCTTTATA	809
Db	9039	AAAGTTATAGAAAACTACATGTAATAAATCTTGATAGCTTAATAAATCTGCCATTTCTTTATA	9098
QY	810	TGGAACGCGATTTGGGTGTTTAAAAAATTTATACAGTTTATAAAGAAAGAAATTTATAAGG	869
Db	9099	TGGAACGCGATTTGGGTGTTTAAAAAATTTATACAGTTTATAAAGAAAGAAATTTGTAAACTA	9158
QY	870	AA 871	
Db	9159	AA 9160	
RESULT 13			
AX538627			
LOCUS	AX538627	12057 bp	DNA linear PAT 23-NOV-2002
DEFINITION	Sequence 47 from Patent WO0229056.		
ACCESSION	AX538627		
VERSION	AX538627.1	GI:25271181	
KEYWORDS	synthetic construct		
SOURCE	synthetic construct		
ORGANISM	artificial sequences.		
REFERENCE	1.		
AUTHORS	Chamberlain, J.S. and Harper, S.O.		
TITLE	Mini-dystrophin nucleic acid and peptide sequences		
JOURNAL	Patent: WO 0229056-A 47 11-APR-2002;		
	THE REGENTS OF THE UNIVERSITY OF MICHIGAN (US)		
FEATURES	Location/Qualifiers		
source	1..12057		
	/organism="synthetic construct"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:32630"		
	/note="Synthetic"		
ORIGIN			
Query Match	41.2%;	Score 410.8;	DB 6; Length 12057;
Best Local Similarity	98.3%;	Pred. NO. 2.2e-73;	
Matches	415;	Conservative	0; Mismatches 7; Indels 0; Gaps 0;
QY	450	AGGACACAATGTAGGAAGTCTTTCCACATGGCAGATGATTTGGGCAGACGATGGAGTC	509
Db	11214	AGGACACAATGTAGGAAGTCTTTCCACATGGCAGATGATTTGGGCAGACGATGGAGTC	11273
QY	510	CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAAATAAATGTTTTACAATCTCTGATT	569
Db	11274	CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAAATAAATGTTTTACAATCTCTGATT	11333
QY	570	CCCCATGCTTTTATTAATTCATACACAAAGAGGATTTAGACAGTAAGAGTTTACAAG	629
Db	11334	CCCCATGCTTTTATTAATTCATACACAAAGAGGATTTAGACAGTAAGAGTTTACAAG	11393
QY	630	AAATAAATCTATATTTTGTGAAGGGPAGTGGTATTTATCTGTAGATTTTCAGTAGTTTCT	689
Db	11394	AAATAAATCTATATTTTGTGAAGGGPAGTGGTATTTATCTGTAGATTTTCAGTAGTTTCT	11453
QY	690	AAGTCTGTTATGTTTTGTTTAAACAATGGCAGGTTTTACAGCTCTATGCAATTTGTACAAA	749
Db	11454	AAGTCTGTTATGTTTTGTTTAAACAATGGCAGGTTTTACAGCTCTATGCAATTTGTACAAA	11513
QY	750	AAGTTTATAGAAAACTACATGTAATAAATCTTGATAGCTTAATAAATCTGCCATTTCTTTATA	809
Db	11514	AAGTTTATAGAAAACTACATGTAATAAATCTTGATAGCTTAATAAATCTGCCATTTCTTTATA	11573
QY	810	TGGAACGCGATTTGGGTGTTTAAAAAATTTATACAGTTTATAAAGAAAGAAATTTATAAGG	869
Db	11574	TGGAACGCGATTTGGGTGTTTAAAAAATTTATACAGTTTATAAAGAAAGAAATTTGTAAACTA	11633
QY	870	AA 871	
Db	11634	AA 11635	

[illegible]

RESULT 15	AX409637	AX409637	13957 bp	DNA	linear	PAT 14-JUN-2002
LOCUS	Sequence 2284	from Patent WO229103.				
DEFINITION	AX409637					
ACCESSION	AX409637					
KEYWORDS	AX409637.1	GI:21442342				
SOURCE	Homo sapiens	(human)				
ORGANISM	Homo sapiens					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
REFERENCE	1	Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.				
AUTHORS		Gene expression profiles in liver cancer				
TITLE	Patent WO 0229103-A	2284 11-APR-2002;				
JOURNAL	GENE LOGIC INC (US)					
FEATURES	Location/Qualifiers					
source	1..13957					
	/organism="Homo sapiens"					
	/mol_type="unassigned DNA"					
	/db_xref="taxon:9606"					
	/note="EMBL/GenBank Accession No. M18533"					
ORIGIN						
	Query Match	41.2%;	Score 410.8;	DB 6;	Length 13957;	
	Best Local Similarity	98.3%;	Pred. No. 2.2e-73;			
	Matches	415;	Conservative	0;	Mismatches	7;
					Indels	0;
					Gaps	0;
Qy	450	AGGACACAAATGTAGGAAGTCTTTCCACATGCGCAGATGATTTGGCGAGCGATGGAGTC	509			
Db	11253	AGGACACAAATGTAGGAAGTCTTTCCACATGCGCAGATGATTTGGCGAGCGATGGAGTC	11312			
Qy	510	CTTAGTATCAGTCATGACAGATGAAAGAGGAGCGAGAATAAATGTTTTCAACTCCTGATT	569			
Db	11313	CTTAGTATCAGTCATGACAGATGAAAGAGGAGCGAGAATAAATGTTTTCAACTCCTGATT	11372			
Qy	570	CCCGCATGTTTTTATAATTATCATACACAAAGAGGATTAGACAGTAAGAGTTTACAAG	629			
Db	11373	CCCGCATGTTTTTATAATTATCATACACAAAGAGGATTAGACAGTAAGAGTTTACAAG	11432			
Qy	630	AAATAAATCTATATTTTGTGAAGGATAGTGATTTATATCTGTAGATTTCAGTAGTTTCT	689			
Db	11433	AAATAAATCTATATTTTGTGAAGGATAGTGATTTATATCTGTAGATTTCAGTAGTTTCT	11492			
Qy	690	AGTCTGTTATGTTTTGTTTACAAATGCCAGGTTTTTACACGCTCTATGCAATTGTACAAA	749			
Db	11493	AGTCTGTTATGTTTTGTTTACAAATGCCAGGTTTTTACACGCTCTATGCAATTGTACAAA	11552			
Qy	750	AAGTTTAAAGAAATACATGATAAAATCTGTAGCTAAATAACTTGCCATTTCTTTATA	809			
Db	11553	AAGTTTAAAGAAATACATGATAAAATCTGTAGCTAAATAACTTGCCATTTCTTTATA	11612			
Qy	810	TGGAACCCATTTGGGTTGTTTTTAAAAAATTATAAACAGTTNTAAAGAAATATAAGG	869			
Db	11613	TGGAACCCATTTGGGTTGTTTTTAAAAAATTATAAACAGTTNTAAAGAAATATAAGG	11672			
Qy	870	AA 871				
Db	11673	AA 11674				

Search completed: April 6, 2004, 13:02:08

Job time : 4177.64 secs

Search completed: April 6, 2004, 13:02:08  
Job time : 4177.64 secs

ORIGIN	Query Match Best Local Similarity Matches 415; Conservative	41.2%; Score 410.8; DB 9; Length 12446; 98.3%; P-red. No. 2.2e-73; 0; Mismatches 7; Indels 0; Gaps 0;
QY	450	AGGACCAATGTAGGAAGTCTTTTCCATGTCAGATGATTTGGGACAGCGATGGAGTC 509
Db	11143	AGGACCAATGTAGGAAGTCTTTTCCATGTCAGATGATTTGGGACAGCGATGGAGTC 11202
QY	510	CTTAGTATCAGTCATGACAGATGGAAGAGGACAGATAAATGTTTTACAACTCCTGATT 569
Db	11203	CTTAGTATCAGTCATGACAGATGGAAGAGGACAGATAAATGTTTTACAACTCCTGATT 11262
QY	570	CCGCGATGTTTTTATAATATTATACACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
Db	11263	CCGCGATGTTTTTATAATATTATACACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11322
QY	630	AAATAAATCTATATTTTTGTGAAGGTTAGTGGTATTATATCTGTAGATTTCAGTAGTTTCT 689
Db	11323	AAATAAATCTATATTTTTGTGAAGGTTAGTGGTATTATATCTGTAGATTTCAGTAGTTTCT 11382
QY	690	AAGTCGTGTTATGTTTTGTATAAATGTAACGAGGTTTACACGCTCATGCAATGGTACAAA 749
Db	11383	AAGTCGTGTTATGTTTTGTATAAATGTAACGAGGTTTACACGCTCATGCAATGGTACAAA 11442
QY	750	AAGTTATAAGAAACTACATGTAAAAATCTTGATAGCTAAATAAATGCGCATTTCTTTATA 809
Db	11443	AAGTTATAAGAAACTACATGTAAAAATCTTGATAGCTAAATAAATGCGCATTTCTTTATA 11502
QY	810	TGGAACGCATTTTGGGTGTTTTAAAAATTTATAACAGTTTATAAGAAAGCAATATATAAGG 869
Db	11503	TGGAACGCATTTTGGGTGTTTTAAAAATTTATAACAGTTTATAAGAAAGCAATATATAACTA 11562
QY	870	AA 871
Db	11563	AA 11564

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: April 6, 2004, 02:56:20 ; Search time 471.188 Seconds  
(without alignment)  
8979.866 Million cell updates/sec

Title: US-09-966-264d-2  
Perfect score: 996  
Sequence: 1 gtggttgatgatagtaaa.....gtgttgatgtaattaatt 996

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 3373863 seqs, 2124099041 residues  
Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_29Jan04: \*  
1: Geneseqn1980s: \*  
2: Geneseqn1990s: \*  
3: Geneseqn2000s: \*  
4: Geneseqn2001as: \*  
5: Geneseqn2001bs: \*  
6: Geneseqn2002s: \*  
7: Geneseqn2003as: \*  
8: Geneseqn2003bs: \*  
9: Geneseqn2003cs: \*  
10: Geneseqn2004s: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	996	100.0	996	6	ABK86497 Human Apo
2	996	100.0	1230	6	ABK86482 Human Apo
3	996	100.0	1234	6	ABK86463 Human Apo
4	435	43.7	1044	6	ABK86464 Mouse gen
5	412.2	41.4	2463	7	ABT33375 NOVX DNA
6	410.8	41.2	5339	6	ABK81998 DNA encod
7	410.8	41.2	5417	6	ABK81997 DNA encod
8	410.8	41.2	5462	6	ABK81999 DNA encod
9	410.8	41.2	8689	6	ABK82000 DNA encod
10	410.8	41.2	11443	6	ABK82002 DNA encod
11	410.8	41.2	12923	1	AAN90338 Sequence
12	410.8	41.2	13957	6	ABK81959 cDNA enco
13	410.8	41.2	13957	6	ABT10904 Human bre
14	410.8	41.2	13957	6	ABN95786 Gene #228
15	410.8	41.2	13957	6	ABN95786 Gene #228
16	409.2	41.1	13977	6	ABK81959 cDNA enco
17	396.8	39.8	2691	6	ABK81996 Human dys
18	374.2	37.6	13815	2	AAV18885 Mus muscu
19	374.2	37.6	13815	6	ABK81960 cDNA enco
20	374.2	37.6	13815	6	ABK81960 cDNA enco
21	374.2	37.6	13815	6	ABK81960 cDNA enco
22	260	26.1	3163	3	AAZ48571 A rod sho
23	260	26.1	3172	3	AAZ48570 A rod sho

24	260	26.1	4075	3	AAZ48569 A rod sho
25	260	26.1	4402	3	AAZ48568 A rod sho
26	260	26.1	4402	3	AAZ48567 A rod sho
27	252.8	25.4	3747	3	AAZ48566 A rod sho
C 28	158	15.9	158	6	ABK86469 Human apo
C 29	151.4	15.2	455	9	ABD32504 Human mit
C 30	151.4	15.2	2691	6	ABK81996 Human dys
C 31	151.4	15.2	8689	6	ABK82000 DNA encod
C 32	151.4	15.2	11443	6	ABK82002 DNA encod
C 33	151.4	15.2	13957	6	ABK81959 cDNA enco
C 34	151.4	15.2	13957	6	ABT10904 Human bre
C 35	151.4	15.2	13957	6	ABN95786 Gene #228
C 36	151.4	15.2	13957	6	ABN95786 Gene #228
C 37	146.4	14.7	2005	7	ABK86468 Human dys
C 38	146.4	14.7	2005	7	ABT33376 NOVX DNA
C 39	140.4	14.1	13977	6	ABK82000 DNA encod
C 40	137	13.8	137	6	ABK86496 Human Apo
C 41	108.4	10.9	959	6	ABN74601 Bovine em
C 42	101.4	10.2	13815	2	AAV18885 Mus muscu
C 43	101.4	10.2	13815	6	ABK81960 cDNA enco
C 44	101.4	10.2	13815	6	ABK81960 cDNA enco
C 45	101.4	10.2	19307	2	ABT27558 Mouse isc
					AAZ48571 A rod sho
					AAZ48570 A rod sho

ALIGNMENTS

RESULT 1  
ABK86497  
ID ABK86497 standard; cDNA; 996 BP.  
XX  
AC ABK86497;  
XX  
DT 27-AUG-2002 (first entry)  
XX  
DE Human Apo-dystrophin-4 cDNA.  
XX  
KW Human; ss; gene; apo-dystrophin-4; inversion sequence; gene therapy;  
KW protein truncation; muscular dystrophy; leukaemia.  
XX  
OS Homo sapiens.

Location/Qualifiers  
1..996  
/tag= a  
/product= "Apo-dystrophin-4 protein appearing as  
AAU98739"  
/partial  
/note= "No start or stop codon shown"  
/transl\_except= (pos:7..9,aa:Xaa)  
/transl\_except= (pos:49..51,aa:Xaa)  
/transl\_except= (pos:187..189,aa:Xaa)  
/transl\_except= (pos:202..204,aa:Xaa)  
/transl\_except= (pos:217..219,aa:Xaa)  
/transl\_except= (pos:253..255,aa:Xaa)  
/transl\_except= (pos:265..270,aa:Xaa-Xaa)  
/transl\_except= (pos:394..396,aa:Xaa)  
/transl\_except= (pos:547..549,aa:Xaa)  
/transl\_except= (pos:565..567,aa:Xaa)  
/transl\_except= (pos:616..618,aa:Xaa)  
/transl\_except= (pos:649..651,aa:Xaa)  
/transl\_except= (pos:682..684,aa:Xaa)  
/transl\_except= (pos:709..711,aa:Xaa)  
/transl\_except= (pos:792..794,aa:Xaa)  
/transl\_except= (pos:841..843,aa:Xaa)  
/transl\_except= (pos:850..852,aa:Xaa)  
/transl\_except= (pos:880..882,aa:Xaa)  
/transl\_except= (pos:907..909,aa:Xaa)  
/transl\_except= (pos:952..954,aa:Xaa)  
/transl\_except= (pos:982..984,aa:Xaa)  
/transl\_except= (pos:988..990,aa:Xaa)  
/note= "Xaa= unknown, encoded by in frame stop codon"

PN GB2368064-A.  
XX 24-APR-2002.  
PD 16-JAN-2001; 2001GB-00001124.  
XX 30-SEP-2000; 2000US-0237079P.  
PR (IMCR ) IMPERIAL CANCER RES TECHNOLOGY LTD.  
PA (BARB/) BARBER E.  
XX Barber E;  
PI WPI; 2002-429042/46.  
DR P-P8DB; AAU98739.  
XX New human regulatory polynucleotide, useful for treating disorders  
PT associated with protein truncation, particularly muscular dystrophy, and  
PT related peptides and antibodies.  
XX Claim 5; Page 170-172; 222pp; English.  
XX The invention relates to a polynucleotide (I) comprising, or consisting  
CC of, apo-dystrophin-4 inversion sequence appearing as ABK86496, or its  
CC functional equivalents (e.g. the apo-dystrophin-4 cDNA sequence appearing  
CC as ABK86497). Also included are polynucleotides that hybridise to either  
CC strand of (I), a vector containing (I), a cell containing (I) or the  
CC vector, proteins and peptides encoded by (I), a protein homologous with  
CC human dystrophin that is expressed on cell surfaces in vivo antibodies  
CC (Ab) specific for the protein and method of screening for leukemia cells  
CC by analysing DNA for presence of (I) or by detecting presence of (II).  
CC The apo-dystrophin-4 inversion sequence is a regulatory element that  
CC controls expression (transcription and translation) of associated DNA,  
CC and may allow read-through of stop codons. The apo-dystrophin-4 inversion  
CC sequence is used in gene therapy of diseases associated with truncation  
CC of proteins, particularly muscular dystrophy and also leukaemia, but more  
CC generally (I) is a regulatory sequence used to control expression of any  
CC attached gene. Analysis of DNA for (I), or detection of proteins (II)  
CC encoded by (I), can be used to screen for leukaemic cells and related  
CC diseases. Antibodies raised against (II) can be used therapeutically, to  
CC inhibit (II) activity, also to detect (II) in screening assays. The  
CC present sequence is the cDNA sequence for human apo-dystrophin-4  
CC containing a plurality of stop codons some of which may be read through  
CC due to the presence of (I) in the apo-dystrophin-4 gene  
XX Sequence 996 BP; 334 A; 152 C; 195 G; 315 T; 0 U; 0 Other;

301 AATCAAACTTGTGTGTATACACCTCGATGACAGCTTCTGTGTGTCTTCCACCCAGAAATG 360  
361 GGAATGATTTCCCAATATGCGCAAGAAACAGAGTGTATCTATCTATCTGACCTTTTGA 420  
361 GGAATGATTTCCCAATATGCGCAAGAAACAGAGTGTATCTATCTATCTGACCTTTTGA 420  
421 AAGTCTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 480  
421 AAGTCTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 480  
481 GCAGATGATTTGGCAGAGGATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 540  
481 GCAGATGATTTGGCAGAGGATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 540  
541 GCAGATGATTTGGCAGAGGATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 600  
541 GCAGATGATTTGGCAGAGGATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 600  
601 AAGAGGATTAGACAGTAAAGAGTTTACAAGAAATAAATCTATATTTTGTGAAGGGTAGTG 660  
601 AAGAGGATTAGACAGTAAAGAGTTTACAAGAAATAAATCTATATTTTGTGAAGGGTAGTG 660  
661 GTATTACTGTAGATTTTCTAAGTCTGTATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 720  
661 GTATTACTGTAGATTTTCTAAGTCTGTATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 720  
721 GTTTTACAGCTCTATGCAATTTGACAAAAGTTTAAAGAAACTACATGTAATAATCTTTG 780  
721 GTTTTACAGCTCTATGCAATTTGACAAAAGTTTAAAGAAACTACATGTAATAATCTTTG 780  
781 ATAGCTAAATAAATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 840  
781 ATAGCTAAATAAATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 840  
841 TAACAGTTTATAAGAAAGAAATTTAAAGGAAAGAAAGAAATTAACGCAATGCAAGTGGTG 900  
841 TAACAGTTTATAAGAAAGAAATTTAAAGGAAAGAAAGAAATTAACGCAATGCAAGTGGTG 900  
901 AAGCTGTGACTCAGGTGTCACAAATATCAGGAACACCCCAAAACCAAGTGAGGTAGA 960  
901 AAGCTGTGACTCAGGTGTCACAAATATCAGGAACACCCCAAAACCAAGTGAGGTAGA 960  
961 AATAGCATGAGAGCGGTGTTGATGTTAATTAATT 996  
961 AATAGCATGAGAGCGGTGTTGATGTTAATTAATT 996

RESULT 2  
ABK86462  
ID ABR86462 standard; cDNA; 1230 BP.  
XX AC ABR86462;  
XX DT 27-AUG-2002 (first entry)  
XX DE Human Apo-dystrophin-4 full length cDNA.  
XX KW Human; ss; gene; apo-dystrophin-4; inversion sequence; gene therapy;  
XX protein truncation; muscular dystrophy; leukaemia.  
XX OS Homo sapiens.  
FH Key Location/Qualifiers  
FT CDS 1..1230  
FT /tag= a  
FT /product= "Apo-dystrophin-4"  
FT /partial  
FT /note= "No start or stop codon shown"  
FT /transl\_except= (pos:142..144,aa:Xaa)  
FT /transl\_except= (pos:241..244,aa:Xaa)  
FT /transl\_except= (pos:282..284,aa:Xaa)  
FT /transl\_except= (pos:421..423,aa:Xaa)  
FT /transl\_except= (pos:436..438,aa:Xaa)







PF 16-JAN-2001; 2001GB-00001124.  
 PR 30-SEP-2000; 2000US-0237079P.  
 XX (IMCR ) IMPERIAL CANCER RES TECHNOLOGY LTD.  
 PA (BARB/) BARBER E.  
 XX Barber E;  
 FI  
 XX  
 XX WP1; 2002-429042/46.  
 XX  
 XX New human regulatory polynucleotide, useful for treating disorders  
 PT associated with protein truncation, particularly muscular dystrophy, and  
 PT related peptides and antibodies.  
 XX  
 XX Disclosure; Fig 9; 222pp; English.  
 XX  
 XX The invention relates to a polynucleotide (I) comprising, or consisting  
 CC of, apo-dystrophin-4 inversion sequence appearing as ABK86496, or its  
 CC functional equivalents (e.g. the apo-dystrophin-4 cDNA sequence appearing  
 CC as ABK86497). Also included are polynucleotides that hybridise to either  
 CC strand of (I), a vector containing (I), a cell containing (I) or the  
 CC vector, proteins and peptides encoded by (I), a protein homologous with  
 CC human dystrophin that is expressed on cell surfaces in vivo antibodies  
 CC (Ab) specific for the protein and method of screening for leukemia cells  
 CC by analysing DNA for presence of (I) or by detecting presence of (II).  
 CC The apo-dystrophin-4 inversion sequence is a regulatory element that  
 CC controls expression (transcription and translation) of associated DNA,  
 CC and may allow read-through of stop codons. The apo-dystrophin-4 inversion  
 CC sequence is used in gene therapy of diseases associated with truncation  
 CC of proteins, particularly muscular dystrophy and also leukaemia, but more  
 CC generally (II) is a regulatory sequence used to control expression of any  
 CC attached gene. Analysis of DNA for (I), or detection of proteins (II)  
 CC encoded by (I), can be used to screen for leukaemic cells and related  
 CC diseases. Antibodies raised against (II) can be used therapeutically, to  
 CC inhibit (II) activity, also to detect (II) in screening assays. The  
 CC present sequence is a region of mouse genomic DNA in the Apo-dystrophin-4  
 XX gene region  
 SQ Sequence 1044 BP; 347 A; 177 C; 200 G; 320 T; 0 U; 0 Other;

Query Match 43.7%; Score 435; DB 6; Length 1044;  
 Best Local Similarity 84.7%; Fred. No. 7.1e-87;  
 Matches 577; Conservative 0; Mismatches 80; Indels 24; Gaps 7;

QY 192 TTTTCCATGTAACACACAGTAGCTGATCCAACTTAAGCTGATACATAACAAACAGT 251  
 DB 257 TTTCCCATCAATGACACCATGCTGATCCAGTATTAACTAATCTAAC- ----ACCA 311  
 QY 252 GTAATGGCTTCAATTAAGGCTTGTCTTCTCGAACTGCTGAAATCAAACTT 311  
 DB 312 TGCAATGCTTCATTAACAAGGATTGCTTCTTGCTAGAAATGGGTAAAAA- -CGGACTGT 369  
 QY 312 GTTGTGTACACCTCGATGCAGCTTCTGTGTGCTTCCACCCAGAAATGGGAATGATTT 371  
 DB 370 GGTCTGTATACCTTCAATGCAGCTTATGTTGCTTTTCC- -TGAAATGGTATGACTC 427  
 QY 372 CCCAAATGGCAAGAAACAGAGTATGCTATCTATCTGCACCTTTTGAAAGCTGTCTTT 431  
 DB 428 CCAATAGTGGCAACACAGGGGTACATACT- - ----TGCAACACTTTGTA- - -CTCTT 476  
 QY 432 TCTTTCTCTTTGTTTCCAGACACAATGATAGGAGTCTTTCCACATGGCAGATGATTT 491  
 DB 477 TCTTTCTCTTTGTTTCCAGACACAATGATAGGAGTCTTTCCACATGGCAGATGATTT 536  
 QY 492 GGGCAGAGCGATGGAGTCTTTAGTATCATGATCATGACAGATGAAGAGAGAGATTAAT 551  
 DB 537 GGGCAGAGCGATGGAGTCTTTAGTATCATGATCATGACAGATGAAGAGAGAGATTAAT 596  
 QY 552 GTTTTACAACTCTGATCCCGCATGGTTTATATATTTTATATATTTTATATATTTTATAT 611  
 DB 597 GTTTTACAACTCTGATCCCGCATGGTTTATATATTTTATATATTTTATATATTTTATAT 656

QY 612 ACAGTAGAGTTTACAGAAAT-AAATCTATATTTTGTGAGGGTAGTGGTATATATACT 670  
 DB 657 ACAGTAGAGTTTACAGAAATATAATTTTGTGAGGGTAGTGGTACTATATACT 716  
 QY 671 GTAGATTTTCAAGTAGTTTCTAAGTCTGTTATTTGTTTAAACAATGGCAGGTTTACACG 730  
 DB 717 GTAGATTTTCAAGTAGTTTCTAAGTCTGTTATTTGTTTAAACAATGGCAGGTTTACACG 776  
 QY 731 TCTATGCAATTTGTACAAAAAGTTTATAGAAAACATCATATTAATCTTGATAGCTAAT 790  
 DB 777 TCTATGCAATTTGTACAAAAAGTTTAAAGAAA- - -ACATGTAATAATCTTGATAGCTAAT 833  
 QY 791 AACTTGGCAATTTCTTATATGGAACGCAATTTTGGGTGTTTAAAAATTTATAACAGTTAT 850  
 DB 834 AACTTGGCAATTTCTTATATGGAACGCAATTTGGGTGTTTAAAAATTTATAACAGTTAT 893  
 QY 851 AAAGAAAGAAATTTATAAGGAA 871  
 DB 894 AAAGAAAGAAATTTATAAGTAAA 914  
 RESULT 5  
 ABT33375  
 ID ABT33375 standard; DNA; 2463 BP.  
 XX  
 AC ABT33375;  
 XX  
 DT 22-MAY-2003 (first entry)  
 XX  
 DE NOVX DNA sequence SEQ ID No 64.  
 XX  
 KW Hepatotropic; immunosuppressive; cardiant; hypertensive; tranquilizer;  
 KW vulnary; viricide; antibacterial; protozoicide; fungicide; nootropic;  
 KW antiparasitic; neuroprotective; cerebroprotective; antiparkinsonian;  
 KW anticonvulsant; antidiabetic; analgesic; dermatological; keratolytic;  
 KW antiseborrheic; antihematuric; antiarthritic; antiinflammatory; anti-HIV;  
 KW cytotoxic; antischismatic; antipsoriatic; hypotensive; osteopathic;  
 KW antitumor; anorectic; antidiabetic; antiallergic; haemostatic;  
 KW neuroleptic; antidepressant; antiinfertility; NOVX; human disease;  
 KW NOVX-associated disorder; trauma; viral; bacterial; fungal; protozoal;  
 KW parasitic infection; Alzheimer's disease; stroke; forensic biology;  
 KW immunogen; non-human transgenic animal; gene therapy; gene; ds.  
 XX Unidentified.  
 XX  
 PN WO200281517-A2.  
 XX  
 PD 17-OCT-2002.  
 XX  
 XX 22-JAN-2002; 2002WO-US002064.  
 PF  
 XX 19-JAN-2001; 2001US-0262892P.  
 PR 23-JAN-2001; 2001US-0263598P.  
 PR 24-JAN-2001; 2001US-0263799P.  
 PR 25-JAN-2001; 2001US-026417P.  
 PR 26-JAN-2001; 2001US-0264139P.  
 PR 26-JAN-2001; 2001US-0264478P.  
 PR 30-JAN-2001; 2001US-0263351P.  
 PR 02-MAR-2001; 2001US-0272870P.  
 PR 14-MAR-2001; 2001US-0275527P.  
 PR 15-MAR-2001; 2001US-0275590P.  
 PR 15-MAR-2001; 2001US-0276449P.  
 PR 20-MAR-2001; 2001US-0277358P.  
 PR 23-MAR-2001; 2001US-0278151P.  
 PR 23-MAR-2001; 2001US-0279857P.  
 PR 20-APR-2001; 2001US-0285140P.  
 PR 20-APR-2001; 2001US-0285141P.  
 PR 30-APR-2001; 2001US-0287484P.  
 PR 17-MAY-2001; 2001US-0291701P.  
 PR 08-JUN-2001; 2001US-0296960P.  
 PR 10-JUL-2001; 2001US-0304353P.  
 PR 10-JUL-2001; 2001US-0304355P.  
 PR 12-JUL-2001; 2001US-0304886P.

PR	09-AUG-2001; 2001US-0311289P.	
PR	13-AUG-2001; 2001US-0311975P.	
PR	16-AUG-2001; 2001US-031237P.	
PR	18-OCT-2001; 2001US-0330227P.	
PR	29-NOV-2001; 2001US-0334198P.	
XX	(CURA-) CURAGEN CORP.	
PA	Decristofaro MF, Padigaru M, Miller C, Tchernev V, Zhong H;	
PI	Zhong M, Anderson D, Ballinger R, Gerlach V, Spytek KA, Rastelli L;	
PI	Kheda R, Guo X, Zerhusen B, Andrew D, Mezes P, Patturajan M;	
PI	Burgess CE, Eisen A, Wolenc A, Baumgartner J, Shinkets RA, Gusev V;	
PI	Vernet CAM, Taupier RJ, Pena C, Shenoy S, Li L, Casman S, Boldog F;	
PI	Fernandes E, Smithson G, Malyankar U, Tallon B, Liu X;	
XX	WPI; 2003-058504/05.	
DR	P-PSDB; AB337309.	
XX	New polypeptides, designated as NOVX, useful for diagnosing and treating	
PT	infections, neurological diseases, cancer, allergy, and bone,	
PT	immunological, skin, renal, brain, muscle and autoimmune disorders.	
XX	Claim 9; Page 171; 672pp; English.	
PS	The invention relates to a novel isolated polypeptide, designated NOVX	
XX	(NOV1 - 33), consisting of a mature form of one of 61 sequences, given in	
CC	the specification, or its variant, where amino acid residue(s) in the	
CC	variant differ from the mature form, provided that the variant differs in	
CC	not more than 15 % of the amino acids from the sequence of the mature	
CC	form. The NOVX polypeptides, nucleic acids encoding the polypeptides, and	
CC	an antibody to the polypeptides, are useful for treating or preventing a	
CC	NOVX-associated disorder in humans and for treating a syndrome associated	
CC	with a human disease (NOVX-associated disorder). NOVX polypeptides and	
CC	the encoding nucleic acids, are useful for determining the presence of or	
CC	predisposition to a disease associated with altered levels of NOVX	
CC	polypeptide and polynucleotide, by measuring the level of polypeptide	
CC	expression or the amount of nucleic acid from a mammal and comparing it	
CC	with another mammal not having or not predisposed to the disease. NOVX	
CC	polypeptide is also useful for identifying an agent that binds to NOVX	
CC	and a cell expressing NOVX is useful for identifying an agent that	
CC	modulates the expression or activity of NOVX. The antibodies and a	
CC	polypeptide having 95 % sequence identity to NOVX polypeptide are useful	
CC	for treating a pathological state in a mammal. The antibodies are also	
CC	useful for determining the presence or amount of NOVX in a sample. NOVX	
CC	polypeptides, polynucleotides and antibodies specific for the	
CC	polypeptides are useful for treating or preventing disorders or syndromes	
CC	including trauma, viral, bacterial, fungal, protozoal, and parasitic	
CC	disease or a stroke. The NOVX encoding nucleic acids are useful for	
CC	expressing the NOVX proteins, to detect NOVX mRNA, or a genetic lesion in	
CC	a NOVX gene and to modulate NOVX activity. NOVX sequences are also useful	
CC	for identifying a cell or tissue type in a biological sample, to amplify	
CC	DNA sequences from very small biological samples such as tissues e.g.	
CC	hair or skin or body fluids in forensic biology and as primers and probes	
CC	for use in identifying and/or cloning NOVX homologues in other cell	
CC	types. The NOVX proteins are useful as an immunogen to generate	
CC	antibodies which are useful for diagnostically monitoring protein levels	
CC	and modulating NOVX activity. Cells comprising NOVX nucleic acids are	
CC	useful for producing non-human transgenic animals which are useful for	
CC	studying the function and/or activity of NOVX protein and for identifying	
CC	and/or evaluating modulators of NOVX protein activity. The NOVX nucleic	
CC	acids can be used in gene therapy. This polynucleotide sequence	
XX	represents a NOVX DNA sequence of the invention	
XX	Sequence 2463 BP; 691 A; 599 C; 580 G; 593 T; 0 U; 0 Other;	
XX	Query Match 41.4%; Score 412.2; DB 7; Length 2463;	
XX	Best Local Similarity 97.0%; Pred.No. 9.8e-82;	
XX	Matches 420; Conservative 0; Mismatch 13; Indels 0; Gaps 0;	
QY	439 CTTTGTTCAGGACACATGAGGAGTCTTTTCACATGCGACATGATTTGGCAGA 498	
DB	1998 CCTAGTTCAAGAGGACACATGAGGAGTCTTTTCACATGCGACATGATTTGGCAGA 2057	

QY	499 GCGATGAGTCCCTTTAGTATCATGACAGATGAGGAGGACAGATAAATGTTTAC 558	
DB	2058 GCGATGAGTCCCTTTAGTATCATGACAGATGAGGAGGACAGATAAATGTTTAC 2117	
QY	559 AACTCCTGATCCCGATCGTTTATATATATATATATATATATATATATATATAT 618	
DB	2118 AACTCCTGATCCCGATCGTTTATATATATATATATATATATATATATATATAT 2177	
QY	619 GAGTTTACAAAGAAATAAATCTATATTTTGTGAAGGAGTGTGTTATATATATAT 678	
DB	2178 GAGTTTACAAAGAAATAAATCTATATTTTGTGAAGGAGTGTGTTATATATATAT 2237	
QY	679 CAGTAGTTTCTAAGTCTGTTATTTGTTTGAAGGAGTGTGTTATATATATATATAT 738	
DB	2238 CAGTAGTTTCTAAGTCTGTTATTTGTTTGAAGGAGTGTGTTATATATATATATAT 2297	
QY	739 ATTGTCAAAAAGTTTATATAGAAAACATCATGTAAATCTTGATAGCTAAATAAAT 798	
DB	2298 ATTGTCAAAAAGTTTATATAGAAAACATCATGTAAATCTTGATAGCTAAATAAAT 2357	
QY	799 ATTCTTTATATGGAACGCAATTTGGTGTGTTTAAATAATTTATACAGTTTAAAGAA 858	
DB	2358 ATTCTTTATATGGAACGCAATTTGGTGTGTTTAAATAATTTATACAGTTTAAAGAA 2417	
QY	859 AATTATAAAGGAA 871	
DB	2418 ATTGTAAACTTAAA 2430	
RESULT 6		
ABK81998		
ID	ABK81998 standard; DNA; 5339 BP.	
XX	AC ABK81998;	
DT	13-AUG-2002 (first entry)	
XX	DNA encoding mini-dystrophin protein deltaR2-R21.	
XX	Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;	
XX	Duchenne's muscular dystrophy; DMD; dystrophin; ds.	
OS	Homo sapiens.	
XX	Synthetic.	
XX	WO200229056-A2.	
XX	11-APR-2002.	
XX	04-OCT-2001; 2001WO-US031126.	
XX	06-OCT-2000; 2000US-0238848P.	
XX	(UNWI ) UNIV MICHIGAN.	
XX	Chamberlain JS, Harper SQ;	
XX	WPI; 2002-435334/46.	
XX	A composition for preparing therapeutic drugs, has a mini-dystrophin	
XX	peptide comprising a specific number of spectrin-like repeat domains, or	
XX	a nucleic acid sequence encoding the mini-dystrophin peptide.	
XX	Example 6; Fig 13; 145pp; English.	
XX	The invention describes a composition comprising a mini-dystrophin	
XX	peptide comprising a spectrin-like repeat domain, where the domain	
XX	comprises n spectrin-like repeats, and contains no more than n spectrin-	
XX	like repeats, where n is an even number between 4-24, or a nucleic acid	
XX	encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the	
XX	polynucleotide encoding it is useful as a medicament, for preparing a	
XX	drug for therapeutic application and in the preparation of a composition	

CC for treatment of muscle disease, e.g. Duchenne's muscular dystrophy  
 CC (DMD). This sequence represents a mini-dystrophin sequence of the  
 CC invention  
 XX  
 SQ Sequence 5339 BP; 1638 A; 1191 C; 1187 G; 1323 T; 0 U; 0 Other;  
 Query Match 41.2%; Score 410.8; DB 6; Length 5339;  
 Best Local Similarity 98.3%; Pred. No. 2.4e-81;  
 Matches 415; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 QY 450 AGGACACAAATGAGGAGTCTTTCCACATGGCAGATGATTTGGCAGAGCGATGGAGTC 509  
 DB 4491 AGGACACAAATGAGGAGTCTTTCCACATGGCAGATGATTTGGCAGAGCGATGGAGTC 4550  
 QY 510 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGATAAATGTTTACAACTCCTGATT 569  
 DB 4551 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGATAAATGTTTACAACTCCTGATT 4610  
 QY 570 CCGCATGGTTTTTATAATTTATATATATATATATATATATATATATATATATATAT 629  
 DB 4611 CCGCATGGTTTTTATAATTTATATATATATATATATATATATATATATATATATAT 4670  
 QY 630 AAATAAATCTATATTTTGTGGAAGGTAGTGTATTTACTGTAGATTTTCAGTAGTTTCT 689  
 DB 4671 AAATAAATCTATATTTTGTGGAAGGTAGTGTATTTACTGTAGATTTTCAGTAGTTTCT 4730  
 QY 690 AAGTCTGTTATTTGTTTGAATGAGGAGTCTTTACACGCTCTATGCAATTTGTACAAA 749  
 DB 4731 AAGTCTGTTATTTGTTTGAATGAGGAGTCTTTACACGCTCTATGCAATTTGTACAAA 4790  
 QY 750 AAGTTATAAGAAAACCTACATGTAATCTTGTAGCTAAATAACTTGGCATTTCTTTATA 809  
 DB 4791 AAGTTATAAGAAAACCTACATGTAATCTTGTAGCTAAATAACTTGGCATTTCTTTATA 4850  
 QY 810 TCGAACGCAATTTGGTGTGTTTAAATTTATAAAGATTTATAAAGAAAGATTTATAAAGG 869  
 DB 4851 TCGAACGCAATTTGGTGTGTTTAAATTTATAAAGATTTATAAAGAAAGATTTATAAAGG 4910  
 QY 870 AA 871  
 DB 4911 AA 4912  
 RESULT 7  
 ABK81997  
 ID ABK81997 standard; DNA; 5417 BP.  
 AC  
 XX  
 AC  
 XX  
 DT 13-AUG-2002 (first entry)  
 XX  
 DE DNA encoding mini-dystrophin protein deltaR4-R23.  
 XX  
 KW Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;  
 KW Duchenne's muscular dystrophy; DMD; dystrophin; ds.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 PN WO200229056-A2.  
 XX  
 PD 11-APR-2002.  
 XX  
 PF 04-OCT-2001; 2001WO-US031126.  
 XX  
 PR 05-OCT-2000; 2000US-023848P.  
 XX  
 PA (UNMI ) UNIV MICHIGAN.  
 XX  
 PI Chamberlain JS, Harper SQ;  
 XX  
 DR WPI; 2002-435334/46.  
 XX

PT A composition for preparing therapeutic drugs, has a mini-dystrophin  
 PT peptide comprising a specific number of spectrin-like repeat domains, or  
 PT a nucleic acid sequence encoding the mini-dystrophin peptide.  
 XX  
 PS Disclosure; Fig 12; 145pp; English.  
 XX  
 CC The invention describes a composition comprising a mini-dystrophin  
 CC peptide comprising a spectrin-like repeat domain, where the domain  
 CC comprises n spectrin-like repeats, and contains no more than n spectrin-  
 CC like repeats, where n is an even number between 4-24, or a nucleic acid  
 CC encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the  
 CC polynucleotide encoding it is useful as a medicament, for preparing a  
 CC drug for therapeutic application and in the preparation of a composition  
 CC for treatment of muscle disease, e.g. Duchenne's muscular dystrophy  
 CC (DMD). This sequence represents a mini-dystrophin sequence of the  
 CC invention  
 XX  
 SQ Sequence 5417 BP; 1700 A; 1192 C; 1182 G; 1343 T; 0 U; 0 Other;  
 Query Match 41.2%; Score 410.8; DB 6; Length 5417;  
 Best Local Similarity 98.3%; Pred. No. 2.4e-81;  
 Matches 415; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 QY 450 AGGACACAAATGAGGAGTCTTTTCCACATGGCAGATGATTTGGCAGAGCGATGGAGTC 509  
 DB 4569 AGGACACAAATGAGGAGTCTTTTCCACATGGCAGATGATTTGGCAGAGCGATGGAGTC 4628  
 QY 510 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGATAAATGTTTACAACTCCTGATT 569  
 DB 4629 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGATAAATGTTTACAACTCCTGATT 4688  
 QY 570 CCGCATGGTTTTTATAATTTATATATATATATATATATATATATATATATATATAT 629  
 DB 4689 CCGCATGGTTTTTATAATTTATATATATATATATATATATATATATATATATATAT 4748  
 QY 630 AAATAAATCTATATTTTGTGGAAGGTAGTGTATTTACTGTAGATTTTCAGTAGTTTCT 689  
 DB 4749 AAATAAATCTATATTTTGTGGAAGGTAGTGTATTTACTGTAGATTTTCAGTAGTTTCT 4808  
 QY 690 AAGTCTGTTATTTGTTTGAATGAGGAGTCTTTACACGCTCTATGCAATTTGTACAAA 749  
 DB 4809 AAGTCTGTTATTTGTTTGAATGAGGAGTCTTTACACGCTCTATGCAATTTGTACAAA 4868  
 QY 750 AAGTTATAAGAAAACCTACATGTAATCTTGTAGCTAAATAACTTGGCATTTCTTTATA 809  
 DB 4869 AAGTTATAAGAAAACCTACATGTAATCTTGTAGCTAAATAACTTGGCATTTCTTTATA 4928  
 QY 810 TCGAACGCAATTTGGTGTGTTTAAATTTATAAAGATTTATAAAGAAAGATTTATAAAGG 869  
 DB 4929 TCGAACGCAATTTGGTGTGTTTAAATTTATAAAGATTTATAAAGAAAGATTTATAAAGG 4988  
 QY 870 AA 871  
 DB 4989 AA 4990  
 RESULT 8  
 ABK81999  
 ID ABK81999 standard; DNA; 5462 BP.  
 AC  
 XX  
 AC  
 XX  
 DT 13-AUG-2002 (first entry)  
 XX  
 DE DNA encoding mini-dystrophin protein deltaR2-R21+H3.  
 XX  
 KW Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;  
 KW Duchenne's muscular dystrophy; DMD; dystrophin; ds.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 PN WO200229056-A2.

XX PD 11-APR-2002.  
 XX PF 04-OCT-2001; 2001WO-US031126.  
 XX PR 06-OCT-2000; 2000US-023884BP.  
 XX PA (UNMI ) UNIV MICHIGAN.  
 XX PI Chamberlain JS, Harper SQ;  
 XX DR WPI; 2002-435334/46.  
 XX PT A composition for preparing therapeutic drugs, has a mini-dystrophin  
 PT peptide comprising a specific number of spectrin-like repeat domains, or  
 PT a nucleic acid sequence encoding the mini-dystrophin peptide.  
 XX PS Disclosure; Fig 14; 145pp; English.  
 XX CC The invention describes a composition comprising a mini-dystrophin  
 CC peptide comprising a spectrin-like repeat domain, where the domain  
 CC comprises n spectrin-like repeats, and contains no more than n spectrin-  
 CC like repeats, where n is an even number between 4-24, or a nucleic acid  
 CC encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the  
 CC polynucleotide encoding it is useful as a medicament, for preparing a  
 CC drug for therapeutic application and in the preparation of a composition  
 CC for treatment of muscle disease, e.g. Duchenne's muscular dystrophy  
 CC (DMD). This sequence represents a mini-dystrophin sequence of the  
 CC invention  
 XX SQ Sequence 5462 BP; 1668 A; 1225 C; 1212 G; 1357 T; 0 U; 0 Other;  
 Query Match 41.2%; Score 410.8; DB 6; Length 5462;  
 Best Local Similarity 98.3%; Pred. No. 2.4e-81;  
 Matches 415; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 QY 450 AGGACACAATGAGGAAGTCTTTCCACATGGCAGATGATTTGGCGAGCGATGGAGTC 509  
 DB 4614 AGGACACAATGAGGAAGTCTTTCCACATGGCAGATGATTTGGCGAGCGATGGAGTC 4673  
 QY 510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAAATAAATGTTTACAACTCCTGATT 569  
 DB 4674 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAAATAAATGTTTACAACTCCTGATT 4733  
 QY 570 CCGCATGTTTTTATAATTCATACAAAGAGGATTAGACAGTAAGATTTTACAAG 629  
 DB 4734 CCGCATGTTTTTATAATTCATACAAAGAGGATTAGACAGTAAGATTTTACAAG 4793  
 QY 630 AAATAAATCTATATTTTGTGAAGGTAGTGGTATTATATCTAGATTTTCAGTAGTTTCT 689  
 DB 4794 AAATAAATCTATATTTTGTGAAGGTAGTGGTATTATATCTAGATTTTCAGTAGTTTCT 4853  
 QY 690 AAGTCGTATTGTTTTGTTTAAACAATGGCAGGTTTTTACACGTCATGCAATTTGTACAAA 749  
 DB 4854 AAGTCGTATTGTTTTGTTTAAACAATGGCAGGTTTTTACACGTCATGCAATTTGTACAAA 4913  
 QY 750 AGTTTATAGAAACTACATGTAATCTTGATAGCTAAATTAATCTTGCCATTTCTTTATA 809  
 DB 4914 AGTTTATAGAAACTACATGTAATCTTGATAGCTAAATCTTGATAGCTAAATCTTGCCATTTCTTTATA 4973  
 QY 810 TGAACGCAATTTGGGTTGTTTAAAAATTTATAACAGTTTATAAAGAAAGATTTATAAAG 869  
 DB 4974 TGAACGCAATTTGGGTTGTTTAAAAATTTATAACAGTTTATAAAGAAAGATTTATAAAG 5033  
 QY 870 AA 871  
 DB 5034 AA 5035  
 RESULT 9  
 ABK82000  
 ID ABK82000 standard; DNA; 8689 BP.  
 XX

ABK82000;  
 13-AUG-2002 (first entry)  
 DNA encoding mini-dystrophin protein deltaH2-R19.  
 Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;  
 Duchenne's muscular dystrophy; DMD; dystrophin; ds.  
 Homo sapiens.  
 Synthetic.  
 WO200229056-A2.  
 11-APR-2002.  
 04-OCT-2001; 2001WO-US031126.  
 06-OCT-2000; 2000US-023884BP.  
 (UNMI ) UNIV MICHIGAN.  
 Chamberlain JS, Harper SQ;  
 WPI; 2002-435334/46.  
 A composition for preparing therapeutic drugs, has a mini-dystrophin  
 peptide comprising a specific number of spectrin-like repeat domains, or  
 a nucleic acid sequence encoding the mini-dystrophin peptide.  
 Disclosure; Fig 15; 145pp; English.  
 The invention describes a composition comprising a mini-dystrophin  
 peptide comprising a spectrin-like repeat domain, where the domain  
 comprises n spectrin-like repeats, and contains no more than n spectrin-  
 like repeats, where n is an even number between 4-24, or a nucleic acid  
 encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the  
 polynucleotide encoding it is useful as a medicament, for preparing a  
 drug for therapeutic application and in the preparation of a composition  
 for treatment of muscle disease, e.g. Duchenne's muscular dystrophy  
 (DMD). This sequence represents a mini-dystrophin sequence of the  
 invention  
 SQ Sequence 8689 BP; 2721 A; 1804 C; 1861 G; 2303 T; 0 U; 0 Other;  
 Query Match 41.2%; Score 410.8; DB 6; Length 8689;  
 Best Local Similarity 98.3%; Pred. No. 2.7e-81;  
 Matches 415; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 QY 450 AGGACACAATGAGGAAGTCTTTCCACATGGCAGATGATTTGGCGAGCGATGGAGTC 509  
 DB 5985 AGGACACAATGAGGAAGTCTTTCCACATGGCAGATGATTTGGCGAGCGATGGAGTC 6044  
 QY 510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAAATAAATGTTTACAACTCCTGATT 569  
 DB 6045 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAAATAAATGTTTACAACTCCTGATT 6104  
 QY 570 CCGCATGTTTTTATAATTCATACAAAGAGGATTAGACAGTAAGATTTTACAAG 629  
 DB 6105 CCGCATGTTTTTATAATTCATACAAAGAGGATTAGACAGTAAGATTTTACAAG 6164  
 QY 630 AAATAAATCTATATTTTGTGAAGGTAGTGGTATTATATCTAGATTTTCAGTAGTTTCT 689  
 DB 6165 AAATAAATCTATATTTTGTGAAGGTAGTGGTATTATATCTAGATTTTCAGTAGTTTCT 6224  
 QY 690 AAGTCGTATTGTTTTGTTTAAACAATGGCAGGTTTTTACAGTCTATGCAATTTGTACAAA 749  
 DB 6225 AAGTCGTATTGTTTTGTTTAAACAATGGCAGGTTTTTACAGTCTATGCAATTTGTACAAA 6284  
 QY 750 AAGTTTATAGAAACTACATGTAATCTTGATAGCTAAATCTTGCCATTTCTTTATA 809  
 DB 6285 AAGTTTATAGAAACTACATGTAATCTTGATAGCTAAATCTTGCCATTTCTTTATA 6344

QY	810	TGGAACGCATTTGGTGTGTTTAAAAATTTATTAACAGTTATTAAGAGAAAGATTTATAAAGG	869
Db	6345	TGGAACGCATTTGGTGTGTTTAAAAATTTATTAACAGTTATTAAGAGAAAGATTTATAAAGG	6404
QY	870	AA 871	
Db	6405	AA 6406	
RESULT 10			
ABK82002			
XX	ABK82002	standard; DNA; 11443 BP.	
XX	AC		
XX	ABK82002;		
XX	13-AUG-2002	(first entry)	
XX			
DE	DNA	encoding mini-dystrophin protein deltaR9-R16.	
XX			
KW	Mini-dystrophin peptide;	spectrin-like repeat domain; muscle disease;	
KM	Duchenne's muscular dystrophy;	DMD; dystrophin; ds.	
XX			
OS	Homo sapiens.		
OS	Synthetic.		
XX			
PN	WO200229056-A2.		
XX			
XX	11-APR-2002.		
XX			
PF	04-OCT-2001;	2001WO-US031126.	
XX			
PR	06-OCT-2000;	2000US-0238848P.	
XX			
PA	(UNMI )	UNIV MICHIGAN.	
XX			
PI	Chamberlain JS,	Harper SQ;	
XX			
DR	WPI;	2002-435334/46.	
XX			
PT	A composition	for preparing therapeutic drugs, has a mini-dystrophin	
PT	peptide	comprising a specific number of spectrin-like repeat domains, or	
PT	a nucleic	acid sequence encoding the mini-dystrophin peptide.	
XX			
PS	Disclosure;	Fig 17; 145pp; English.	
XX			
CC	The invention	describes a composition comprising a mini-dystrophin	
CC	peptide	comprising a spectrin-like repeat domain, where the domain	
CC	comprises	n spectrin-like repeats, and contains no more than n spectrin-	
CC	like repeats,	where n is an even number between 4-24, or a nucleic acid	
CC	encoding	a mini-dystrophin peptide. The mini-dystrophin peptide or the	
CC	polynucleotide	encoding it is useful as a medicament, for preparing a	
CC	drug	for therapeutic application and in the preparation of a composition	
CC	for treatment	of muscle disease, e.g. Duchenne's muscular dystrophy	
CC	(DMD).	This sequence represents a mini-dystrophin sequence of the	
CC	invention		
XX			
SQ	Sequence	11443 BP; 3707 A; 2339 C; 2502 G; 2895 T; 0 U; 0 Other;	
Query Match		41.2%; Score 410.8; DB 6; Length 11443;	
Best Local Similarity		98.3%; Pred. No. 2.8e-81;	
Matches 415;	Conservative	0; Mismatches 7; Indels 0; Gaps 0;	
QY	450	AGGACCAATGTAGGAAGTCTTTTCCACATGCGAGATGATTTGGGCAGAGCGATGGAGTC	509
Db	8739	AGGACCAATGTAGGAAGTCTTTTCCACATGCGAGATGATTTGGGCAGAGCGATGGAGTC	8799
QY	510	CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGATAAATGTTTTCACATCTCTGATT	569
Db	8799	CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGATAAATGTTTTCACATCTCTGATT	8859
QY	570	CCGCGATGTTTTTTTAAATATTCATACAAAGAGGATTAGACAGTAAAGAGTTTACAAG	629
Db	8859	CCGCGATGTTTTTTTAAATATTCATACAAAGAGGATTAGACAGTAAAGAGTTTACAAG	8919



```
XX Orr MS, Nation M, Diggins JC, Zeng W;
XX WPI; 2002-674803/72.
XX Diagnosing breast cancer in a patient comprises detecting the level of
XX gene expression in cell or tissue samples, where a differential gene
XX expression is indicative of breast cancer.
XX
XX Claim 1; SEQ ID NO 1038; 260pp + Sequence Listing; English.
XX
XX The present invention relates to methods of diagnosing breast cancer in a
XX patient, which comprise detecting the level of expression in a tissue
XX sample of two or more genes selected from those shown in AB109867-
XX ABR1112, where a differential expression of the genes indicates breast
XX cancer. The methods are useful in diagnosing, treating, detecting the
XX progression, and in monitoring treatment of breast cancer in patients.
XX The methods are also useful as a screening tool for agents that modulate
XX the onset or progression of breast cancer. The breast cancer genes may be
XX used as diagnostic markers for the prediction or identification of the
XX malignant state of breast tissue, for confirming the type and progression
XX of cancer, and for drug screening and assays. The present sequence is a
XX coding sequence of the invention. Note: The sequence data for this patent
XX did not form part of the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Query Match 41.2%; Score 410.8; DB 6; Length 13957;
XX Best Local Similarity 98.3%; Pred. No. 3e-81;
XX Matches 415; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
XX
XX QY 450 AGGACACAATGTAGGAAGTCTTTCCACATGGCAGATGATTTGGCAGAGCGATGGAGTC 509
XX DB 11253 AGGACACAATGTAGGAAGTCTTTCCACATGGCAGATGATTTGGCAGAGCGATGGAGTC 11312
XX
XX QY 510 CTTAGTATCAGTCATGACAGATGAAGAAGGACAGATAAATGTTTACAACCTCTGATT 569
XX DB 11313 CTTAGTATCAGTCATGACAGATGAAGAAGGACAGATAAATGTTTACAACCTCTGATT 11372
XX
XX QY 570 CCGCATGGTCTTTTATAATTTATATTCATACACAAGAGGATTAGACAGTAAGAGTTTACAAG 629
XX DB 11373 CCGCATGGTCTTTTATAATTTATATTCATACACAAGAGGATTAGACAGTAAGAGTTTACAAG 11432
XX
XX QY 630 AAATAAATCTATATTTTGTGAAGGTTAGTGTGTTATATCTATGATTTTCAGTAGTTTCT 689
XX DB 11433 AAATAAATCTATATTTTGTGAAGGTTAGTGTGTTATATCTATGATTTTCAGTAGTTTCT 11492
XX
XX QY 690 AAGTCTGTTATTTGTTTAAACAATGGCAGGTTTACAGCTCTATGCAATTTGACAAAA 749
XX DB 11493 AAGTCTGTTATTTGTTTAAACAATGGCAGGTTTACAGCTCTATGCAATTTGACAAAA 11552
XX
XX QY 750 AAGTTTAAAGAAAACCTACATGTAATAATCTTGATAGCTAAATTAACCTTGCATTTCTTTATA 809
XX DB 11553 AAGTTTAAAGAAAACCTACATGTAATAATCTTGATAGCTAAATTAACCTTGCATTTCTTTATA 11612
XX
XX QY 810 TCGAACGCAATTTGGTGTGTTTAAATAATTTAATACAGTTTAAAGAAATTTAAGG 869
XX DB 11613 TCGAACGCAATTTGGTGTGTTTAAATAATTTAATACAGTTTAAAGAAATTTAAGG 11672
XX
XX QY 870 AA 871
XX DB 11673 AA 11674
XX
XX RESULT 14
XX ID ABN95786
XX ABN95786 standard; DNA; 13957 BP.
XX
XX AC ABN95786;
XX
XX DT 13-AUG-2002 (first entry)
```

```
XX Gene #2284 used to diagnose liver cancer.
XX
XX Gene: liver cancer; de; hepatocellular carcinoma; hepatotropic;
XX metastatic liver tumour; cytostatic; expression profile; disease state;
XX disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
XX OS Homo sapiens.
XX PN WO200229103-A2.
XX PD 11-APR-2002.
XX PF 02-OCT-2001; 2001WO-US030589.
XX PR 02-OCT-2000; 2000US-0237054P.
XX PA (GENE-) GENE LOGIC INC.
XX PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX WPI; 2002-426119/45.
XX
XX Diagnosing and detecting the progression of liver cancer, hepatocellular
XX carcinoma or metastatic liver tumor in a patient, involves detecting the
XX level of expression of two or more genes in a liver tissue sample.
XX
XX Claim 1; SEQ ID NO 2284; 298pp; English.
XX
XX The invention relates to a novel method for diagnosing and detecting the
XX progression of liver cancer, hepatocellular carcinoma or metastatic liver
XX tumour in a patient, and differentiating metastatic liver cancer from
XX hepatocellular carcinoma in a patient, involving detecting the level of
XX expression of two or more genes represented in ABN93503-ABN97455 in a
XX tissue sample. The method of the invention has hepatotropic, and
XX cytostatic activity. The method is useful for diagnosing and detecting
XX the progression of liver cancer, hepatocellular carcinoma and metastatic
XX liver carcinoma in a patient. The method is useful for identifying
XX expression profiles which serve as useful diagnostic markers as well as
XX markers that can be used to monitor disease states, disease progression,
XX drug toxicity, drug efficacy and drug metabolism. Note: The sequence data
XX for this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 13957 BP; 4602 A; 2781 C; 3122 G; 3452 T; 0 U; 0 Other;
XX
XX Query Match 41.2%; Score 410.8; DB 6; Length 13957;
XX Best Local Similarity 98.3%; Pred. No. 3e-81;
XX Matches 415; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
XX
XX QY 450 AGGACACAATGTAGGAAGTCTTTCCACATGGCAGATGATTTGGCAGAGCGATGGAGTC 509
XX DB 11253 AGGACACAATGTAGGAAGTCTTTCCACATGGCAGATGATTTGGCAGAGCGATGGAGTC 11312
XX
XX QY 510 CTTAGTATCAGTCATGACAGATGAAGAAGGACAGATAAATGTTTACAACCTCTGATT 569
XX DB 11313 CTTAGTATCAGTCATGACAGATGAAGAAGGACAGATAAATGTTTACAACCTCTGATT 11372
XX
XX QY 570 CCGCATGGTCTTTTATAATTTATATTCATACACAAGAGGATTAGACAGTAAGAGTTTACAAG 629
XX DB 11373 CCGCATGGTCTTTTATAATTTATATTCATACACAAGAGGATTAGACAGTAAGAGTTTACAAG 11432
XX
XX QY 630 AAATAAATCTATATTTTGTGAAGGTTAGTGTGTTATATCTATGATTTTCAGTAGTTTCT 689
XX DB 11433 AAATAAATCTATATTTTGTGAAGGTTAGTGTGTTATATCTATGATTTTCAGTAGTTTCT 11492
XX
XX QY 690 AAGTCTGTTATTTGTTTAAACAATGGCAGGTTTACAGCTCTATGCAATTTGACAAAA 749
XX DB 11493 AAGTCTGTTATTTGTTTAAACAATGGCAGGTTTACAGCTCTATGCAATTTGACAAAA 11552
XX
XX QY 750 AAGTTTAAAGAAAACCTACATGTAATAATCTTGATAGCTAAATAAATCTTGCATTTCTTTATA 809
```

Db 11553 AAGTTATAGAAAACACATGTAATAATCTTGATAGCTAAATTAACCTGGCCATTCCTTTATA 11612  
QY 810 TCGAACGCAATTTGGTGTGTTTAAAAATTTAATACAGTTTAAAGAAATTAATAAGG 869  
Db 11613 TCGAACGCAATTTGGTGTGTTTAAAAATTTAATACAGTTTAAAGAAATTTGTAACCTA 11672  
QY 870 AA 871  
Db 11673 AA 11674  
RESULT 15  
ABS69900  
ID ABS69900 standard; DNA; 13957 BP.  
XX  
AC ABS69900;  
XX  
DT 21-NOV-2002 (first entry)  
XX  
DE Human dystrophin gene.  
XX  
KW Vector; adenovirus; adeno-associated; adenosine deaminase gene; receptor;  
KW adenosine deaminase deficiency; severe combined immune deficiency; PAH;  
KW beta-chain; haemoglobin gene; beta-thalassaemia; sickle cell disease;  
KW low density lipoprotein gene; familial hypercholesterolaemia;  
KW hypoxanthine-guanine phosphoribosyltransferase; Lesch-Nyhan syndrome;  
KW phenylalanine hydroxylase gene; gene therapy; phenylketonuria;  
KW dystrophin gene; muscular dystrophy; cystic fibrosis; immunostimulant;  
KW human cystic fibrosis transmembrane conductance regulator gene;  
KW antianemic; antilipemic; noctropic; cycostatic; dermatological; human;  
KW alpha-1-antitrypsin; lysosomal glucocerebrosidase; ADA; HPRT;  
KW lysosomal arylsulphatase A; ornithine transcarbamylase; ARSA; OTC; NP;  
KW purin nucleoside phosphorylase; gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN US2002102731-A1.  
XX  
PD 01-AUG-2002.  
XX  
PF 12-FEB-2001; 2001US-00782378.  
XX  
PR 02-OCT-2000; 2000US-0237747P.  
XX  
PA (JYNY ) UNIV NEW YORK STATE RES FOUND.  
XX  
PI Hearing P, Bahou WF, Sandalon Z, Gnatenko DV;  
XX  
PS WPI; 2002-690619/74.  
XX  
PT Producing vector, by introducing vector having nucleotide sequence,  
PT adenovirus inverted terminal repeats and packaging sequence, and adeno-  
PT associated virus terminal repeat, into cell, and culturing cell.  
XX  
PS Disclosure; Page 122-128; 191pp; English.  
XX  
CC The present invention relates to a new method of producing a vector. The  
CC method involves introducing recombinant vector having nucleotide sequence  
CC (NS) having 5' and 3' end, left and right inverted terminal repeats of  
CC adenovirus flanking NS, adenovirus packaging sequence linked to inverted  
CC terminal repeat, and adeno-associated virus terminal repeat linked to 3'  
CC end of NS, into cell expressing adenovirus early gene lacking from vector  
CC ; and culturing cell to produce another vector. The method is useful for  
CC generating vectors, especially mad vectors. The method is useful in  
CC transferring nucleotide sequences of interest into a cell, for gene  
CC transfer applications (e.g. gene therapy) in vitro, ex vivo and in vivo.  
CC The nucleotide sequences are useful for treating diseases associated with  
CC it, i.e. adenosine deaminase gene associated with adenosine deaminase  
CC deficiency with severe combined immune deficiency, beta-chain of  
CC haemoglobin gene associated with beta-thalassaemia and sickle cell  
CC disease, receptor for low density lipoprotein gene associated with  
CC familial hypercholesterolaemia, hypoxanthine-guanine  
CC phosphoribosyltransferase associated with Lesch-Nyhan syndrome,

CC phenylalanine hydroxylase (PAH) gene associated with phenylketonuria;  
CC dystrophin gene associated with muscular dystrophy, and human cystic  
CC fibrosis transmembrane conductance regulator gene associated with cystic  
CC fibrosis. The present nucleic acid sequence represents a human disease  
CC gene sequence that was used in the methods of the invention  
XX  
SQ Sequence 13957 BP; 4602 A; 2781 C; 3122 G; 3452 T; 0 U; 0 Other;  
Query Match 41.2%; Score 410.8; DB 6; Length 13957;  
Best Local Similarity 98.3%; Pred. NO. 3e-81; 7; Indels 0; Gaps 0;  
Matches 415; Conservative 0; Mismatches 0;  
QY 450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCGAGCGATGGAGTC 509  
Db 11253 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCGAGCGATGGAGTC 11312  
QY 510 CTTAGTATCATGATGACAGATGAAGAGGAGCAGATAAATGTTTTTACAACCTCCTGATT 569  
Db 11313 CTTAGTATCATGATGACAGATGAAGAGGAGCAGATAAATGTTTTTACAACCTCCTGATT 11372  
QY 570 CCGCATGCTGTTTTTATAATATTCATACACAAAGAGGATTTAGACAGTAAGAGTTTACAAG 629  
Db 11373 CCGCATGCTGTTTTTATAATATTCATACACAAAGAGGATTTAGACAGTAAGAGTTTACAAG 11432  
QY 630 AATATAATCTATATTTTGTGAAGGTTAGTGGTATTATCTAGTATGATTTTCAGTAGTTTCT 689  
Db 11433 AATATAATCTATATTTTGTGAAGGTTAGTGGTATTATCTAGTATGATTTTCAGTAGTTTCT 11492  
QY 690 AAGTCTCTTATTTGTTTTTAAACAATGGCAGGTTTTTACACGCTCTATGCAATTTGACAAA 749  
Db 11493 AAGTCTCTTATTTGTTTTTAAACAATGGCAGGTTTTTACACGCTCTATGCAATTTGACAAA 11552  
QY 750 AAGTTATAAGAAAACCTACATGTAATAATCTTTGATAGCTAAATACTTGCCATTTCTTTATA 809  
Db 11553 AAGTTATAAGAAAACCTACATGTAATAATCTTTGATAGCTAAATACTTGCCATTTCTTTATA 11612  
QY 810 TGGACCGCATTTGGGTTGTTTTAAATAATTTATAACAGTTTATAAGAAAGAAATTTATAAGG 869  
Db 11613 TGGACCGCATTTGGGTTGTTTTAAATAATTTATAACAGTTTATAAGAAAGAAATTTATAAGG 11672  
QY 870 AA 871  
Db 11673 AA 11674

Search completed: April 6, 2004, 11:42:39  
Job time : 473.188 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 08:36:01 ; Search time 110.764 Seconds

(without alignments)  
4990.154 Million cell updates/sec

Title: US-09-966-264D-2

Perfect score: 996

Sequence: 1 gtgtttgatgatagtaaa.....gtgtttgatgtaattaatt 996

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Capext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA:\*

1: /cgn2\_6/ptodata/2/ina/5A COMB.seq:\*

2: /cgn2\_6/ptodata/2/ina/5B COMB.seq:\*

3: /cgn2\_6/ptodata/2/ina/6A COMB.seq:\*

4: /cgn2\_6/ptodata/2/ina/6B COMB.seq:\*

5: /cgn2\_6/ptodata/2/ina/PTUS COMB.seq:\*

6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	409.2	41.1	13977	4	US-09-484-970B-60
C 2	374.2	37.6	19307	3	US-08-836-022A-10
C 3	374.2	37.6	19307	3	US-08-836-022A-10
C 4	140.4	14.1	13977	4	US-09-484-970B-60
5	101.4	10.2	19307	3	US-08-836-022A-10
6	101.4	10.2	19307	3	US-09-427-048A-10
C 7	45	4.5	615	3	US-08-998-416-186
8	44	4.4	8093	4	US-10-204-708-31
C 9	43.5	4.4	832	4	US-09-621-976-2813
10	43	4.3	1830121	4	US-09-557-884-1
11	43	4.3	1830121	4	US-09-557-884-1
C 12	42.8	4.3	7218	1	US-08-232-463-14
13	42.4	4.3	19124	2	US-08-487-826B-13
14	42	4.2	5156	2	US-09-091-432-3
15	42	4.2	5156	4	US-09-387-663-3
16	42	4.2	5156	4	US-09-214-139B-3
C 17	41.5	4.2	640681	4	US-09-790-988-1
18	41.2	4.1	11049	4	US-10-204-708-21
19	41	4.1	15567	4	US-09-627-376-3
20	40.6	4.1	615	3	US-08-998-416-186
21	40.6	4.1	636	3	US-08-998-416-1137
22	40.6	4.1	837	3	US-08-998-416-288
C 23	40.4	4.1	5152	4	US-10-204-708-73
C 24	40.4	4.1	6124	4	US-08-213-419B-3
C 25	40.4	4.1	319608	4	US-09-539-333D-1
C 26	40.4	4.1	319608	4	US-09-679-409-1
27	40	4.0	6124	4	US-08-213-419B-3

#### ALIGNMENTS

##### RESULT 1

US-09-484-970B-60

; Sequence 60, Application US/09484970B

; Patent No. 6426186

; GENERAL INFORMATION:

; APPLICANT: Jones, Karen A.

; APPLICANT: Volkmutch, Wayne

; APPLICANT: Walker, Michael G.

; TITLE OF INVENTION: BONE REMODELING GENES

; FILE REFERENCE: PB-0014 US

; CURRENT APPLICATION NUMBER: US/09/484,970B

; CURRENT FILING DATE: 2000-01-18

; NUMBER OF SEQ ID NOS: 172

; SOFTWARE: PERL Program

; SEQ ID NO 60

; LENGTH: 13977

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; OTHER INFORMATION: Incyte ID No. 6426186 229357.11CBI

; NAME/KEY: unsure

; LOCATION: 11721-11761, 12294, 13969

; OTHER INFORMATION: a, t, c, g, or other

US-09-484-970B-60

Query Match Similarity 41.1%; Score 409.2; DB 4; Length 13977;

Best Local Similarity 98.1%; Pred. No. 1.6e-92; Indels 0; Gaps 0;

Matches 414; Conservative 0; Mismatches 8;

QY	450	AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTGGCGAGCGATGGAGTC	509
DB	11254	AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTGGCGAGCGATGGAGTC	11313
QY	510	CTTAGTATCATGTATGACAGATGAAGAGGAGCAATATGTTTACAACTCCTGATT	569
DB	11314	CTTAGTATCATGTATGACAGATGAAGAGGAGCAATATGTTTACAACTCCTGATT	11373
QY	570	CCCGCATGGTTTTTATATATATTCATACAAAGAGGATTAGACAGTAAGAGTTTACAAG	629
DB	11374	CCCGCATGGTTTTTATATATATTCATACAAAGAGGATTAGACAGTAAGAGTTTACAAG	11433
QY	630	AAATAAATCTATATTTTGTGAAGGATGTTGTTTACAGTATGATTTCAGTAGTTCT	689
DB	11434	AAATAAATCTATATTTTGTGAAGGATGTTGTTTACAGTATGATTTCAGTAGTTCT	11493
QY	690	AAGTCTGTTATTTGTTTAAACAATGGCAGGTTTTTACAGCTCTATGCAATTTGACAAA	749
DB	11494	AAGTCTGTTATTTGTTTAAACAATGGCAGGTTTTTACAGCTCTATGCAATTTGACAAA	11553

Sequence 7, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 5, Appli  
Sequence 23, Appli  
Sequence 82, Appli  
Sequence 46, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 2813, Ap  
Sequence 1, Appli  
Patent No. 5256558  
Sequence 6, Appli  
Sequence 3, Appli  
Sequence 58, Appli  
Sequence 1797, Ap  
Sequence 1797, Ap  
Sequence 1797, Ap

28 39.8 4.0 4185 4 US-09-417-485D-7  
29 39.8 4.0 8920 2 US-08-446-855A-1  
30 39.8 4.0 8920 3 US-09-150-741-1  
31 39.8 4.0 10640 4 US-09-417-485D-5  
32 39.6 4.0 17310 4 US-08-956-171E-23  
33 39.4 4.0 1661 3 US-08-821-994-82  
34 39.4 4.0 19233 4 US-10-204-708-46  
35 39.2 3.9 640681 4 US-09-790-988-1  
36 39 3.9 1664976 4 US-08-916-421B-1  
37 38.8 3.9 832 4 US-09-621-976-2813  
38 38.8 3.9 51952 3 US-08-947-823-1  
39 38.4 3.9 1121 6 5256558-7  
40 38.4 3.9 1241 1 US-07-593-657-6  
41 38.4 3.9 1241 3 US-08-942-012B-3  
42 38.4 3.9 3312 4 US-09-601-198-58  
43 38.4 3.9 4600 4 US-09-702-705-1797  
44 38.4 3.9 4600 4 US-09-736-457-1797  
45 38.4 3.9 4600 4 US-09-671-325-1797

QY 750 AAGTTATAGAAAACCTACATGTAATAATCTTGATAGCTAAATAACTTGCCTTCCTTTATA 809  
Db 11554 AAGTTATAGAAAACCTACATGTAATAATCTTGATAGCTAAATAACTTGCCTTCCTTTATA 11613  
QY 810 TCGAACGCAATTTGGTGGTGTATTAATAATTTATTAACAGTTATAAGAAAGAAATTAAGG 869  
Db 11614 TCGAACGCAATTTGGTGGTGTATTAATAATTTATTAACAGTTATAAGAAAGAAATTAAGG 11673  
QY 870 AA 871  
Db 11674 AA 11675

RESULT 2

US-08-836-022A-10/c  
; Sequence 10, Application US/08836022A  
; Patent No. 6001557  
; GENERAL INFORMATION:  
; APPLICANT: Trustees of the University of Pennsylvania  
; APPLICANT: Wilson, James M.  
; APPLICANT: Fisher, Krishna J.  
; APPLICANT: Chen, Shu-Jen  
; APPLICANT: Weitzman, Matthew  
; TITLE OF INVENTION: Improved Adenovirus Virus and  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Howson and Howson  
; STREET: Spring House Corporate Cntr, P O Box 457  
; CITY: Spring House  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19477  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/836,022A  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/331,381  
; FILING DATE: 28-OCT-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bak, Mary E.  
; REGISTRATION NUMBER: 31,215  
; REFERENCE/DOCKET NUMBER: GNVN.008PCT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-540-9200  
; TELEFAX: 215-540-5818  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 19307 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: CDNA  
; US-08-836-022A-10

Query Match 37.6%; Score 374.2; DB 3; Length 19307;  
Best Local Similarity 96.0%; Pred. No. 9.5e-84;  
Matches 406; Conservative 0; Mismatches 13; Indels 4; Gaps 2;  
QY 450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCGAGCGCATGGAGTC 509  
Db 3441 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCGAGCGCATGGAGTC 3382  
QY 510 CTTAGTATCATGTCATGACATGAGAGGAGAGAGATTAATGTTTACAACTCCTGATT 569  
Db 3381 CTTAGTATCATGTCATGACATGAGAGGAGAGAGATTAATGTTTACAACTCCTGATT 3322  
QY 570 CCGCATGGTTTTTATTAATATTTCATACAAAGAGAGATTAGACAGTAAGAGTTTACAAG 629

Db 3321 CCGCATGGTTTTTATTAATATTTCATACAAAGAGAGATTAGACAGTAAGAGTTTACAAG 3262  
QY 630 AAT-AAAATCTATTTTGTGAAGGAGTAGTGTATTATTAATCTAGATTTCAGTAGTTTC 688  
Db 3261 AAT-AAAATCTATTTTGTGAAGGAGTAGTGTATTATTAATCTAGATTTCAGTAGTTTC 3202  
QY 689 TAAGTCTGTATTTTGTGAAGGAGTAGTGTATTATTAATCTAGATTTCAGTAGTTTC 748  
Db 3201 TAAGTCTGTATTTTGTGAAGGAGTAGTGTATTATTAATCTAGATTTCAGTAGTTTC 3142  
QY 749 AAGTTATAGAAAACCTACATGTAATAATCTTGATAGCTAAATAACTTGCCTTCCTTTATA 808  
Db 3141 AAGTTATAGAAAACCTACATGTAATAATCTTGATAGCTAAATAACTTGCCTTCCTTTATA 3085  
QY 809 ATGGAACGCAATTTGGTGGTGTATTAATAATTTATTAACAGTTATAAGAAAGAAATTAAGG 868  
Db 3084 ATGGAACGCAATTTGGTGGTGTATTAATAATTTATTAACAGTTATAAGAAAGAGATTGTAAGT 3025  
QY 869 GAA 871  
Db 3024 AAA 3022

RESULT 3

US-09-427-048A-10/c  
; Sequence 10, Application US/09427048A  
; Patent No. 6203975  
; GENERAL INFORMATION:  
; APPLICANT: Trustees of the University of Pennsylvania  
; APPLICANT: Wilson, James M.  
; APPLICANT: Fisher, Krishna J.  
; APPLICANT: Chen, Shu-Jen  
; APPLICANT: Weitzman, Matthew  
; TITLE OF INVENTION: Improved Adenovirus Virus and  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Howson and Howson  
; STREET: Spring House Corporate Cntr, P O Box 457  
; CITY: Spring House  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19477  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/427,048A  
; FILING DATE: 21-Oct-1999  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/836,022  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bak, Mary E.  
; REGISTRATION NUMBER: 31,215  
; REFERENCE/DOCKET NUMBER: GNVN.008PCT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-540-9200  
; TELEFAX: 215-540-5818  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 19307 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: CDNA  
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:  
US-09-427-048A-10

```
Query Match      37.6%; Score 374.2; DB 3; Length 19307;
Best Local Similarity 96.0%; Pred. No. 9.5e-84;
Matches 406; Conservative 0; Mismatches 13; Indels 4; Gaps 2;

QY 450 AGGACACAAATGAGGAAGTCTTTTCCACATGCGCAGATGATTTGGCAGAGCGATGGAGTC 509
DB 3441 AGGACACAAATGAGGAAGTCTTTTCCACATGCGCAGATGATTTGGCAGAGCGATGGAGTC 3382

QY 510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCGAGATAAATGTTTACAACTCCTGATT 569
DB 3381 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCGAGATAAATGTTTACAACTCCTGATT 3322

QY 570 CCCGATGTTTTTATAATATTCATACAAAGAGGATTAGACAGTAAGAGTTTACAAAG 629
DB 3321 CCCGATGTTTTTATAATATTCGTAACAAAGAGGATTAGACAGTAAGAGTTTACAAAG 3262

QY 630 AAAT-AAATCTATATTTTGTGAAGGAGTGGTATTATATCTAGATTTTTCAGTAGTTTC 688
DB 3261 AAATAAATCTATATTTTGTGAAGGAGTGGTATTATATCTAGATTTTTCAGTAGTTTC 3202

QY 689 TAAGTCTGTTATGTTTGTAAACATGCGAGGTTTACAGTCTATGCAATTTACAAA 748
DB 3201 TAAGTCTGTTATGTTTGTAAACATGCGAGGTTTACAGTCTATGCAATTTACAAA 3142

QY 749 AAAGTTATAAGAAAACATCATGTAATAATCTTGATAGCTAAATAACTTCCCATTTCTTTAT 808
DB 3141 AAAGTTATAAGAAAACATCATGTAATAATCTTGATAGCTAAATAACTTCCCATTTCTTTAT 3085

QY 809 ATGGAACGCAATTTGGGTTGTTTAAATTTTAAACAGTTTAAAGAAAGAAATTAAGA 868
DB 3084 ATGGAACGCAATTTGGGTTGTTTAAATTTTAAACAGTTTAAAGAAAGAAATTTAAACT 3025

QY 869 GAA 871
DB 3024 AAA 3022

RESULT 4
US-09-484-970B-60/c
; Sequence 60, Application US/09484970B
; Patent No. 6426186
; GENERAL INFORMATION:
; APPLICANT: Jones, Karen A.
; APPLICANT: Volkmut, Wayne
; APPLICANT: Walker, Michael G.
; TITLE OF INVENTION: BONE REMODELING GENES
; FILE REFERENCE: PB-0014 US
; CURRENT APPLICATION NUMBER: US/09/484,970B
; CURRENT FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: PERL Program
; SEQ ID NO 60
; LENGTH: 13977
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6426186 229357.11CB1
; NAME/KEY: unsure
; LOCATION: 11721-11761, 13294, 13969
; OTHER INFORMATION: a, t, c, g, or other
US-09-484-970B-60

Query Match      14.1%; Score 140.4; DB 4; Length 13977;
Best Local Similarity 90.4%; Pred. No. 1.8e-25;
Matches 161; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

QY 820 TTGGGTTGTTTAAATTTATTAACAGTTTATAAGAAAGAAATTTATAAGGAAAAAGAAAA 879
DB 13346 TTGGGTTTCTTTTGAATTTATGAGGAAAGAAAGAAATTTATAAGGAAAAAGAAAA 13287

QY 880 TAACGAATGGCAAGTGGTGAAGCTGTGAATCTCAGGTGTGCACATTTATCAGGAACACC 939
```

```
DB 13286 TAACGCAATGGCAAGTGGTGAAGCTGTGAATCTCAGGTGTGCACATTTATCAGGAACACC 13227
QY 940 CCAAAACCAAGTGAAGTGAATAATAGCATGAGAAG-CCGTGTTTGTATGTTAATTAATT 996
DB 13226 CCAAAACCAAGTGAAGTGAATAATAGCATGAGAAGCCCGTGTGATGTTAATTAATT 13169

RESULT 5
US-08-836-022A-10
; Sequence 10, Application US/08836022A
; Patent No. 6001557
; GENERAL INFORMATION:
; APPLICANT: Trustees of the University of Pennsylvania
; APPLICANT: Wilson, James M.
; APPLICANT: Fisher, Krishna J.
; APPLICANT: Chen, Shu-Jen
; APPLICANT: Weitzman, Matthew
; TITLE OF INVENTION: Improved Adenovirus Virus and
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P O Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,022A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/331,381
; FILING DATE: 28-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: GNVPN.008PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19307 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cdna
US-08-836-022A-10

Query Match      10.2%; Score 101.4; DB 3; Length 19307;
Best Local Similarity 81.8%; Pred. No. 1e-15;
Matches 117; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 851 AAGAAAGAAATTTATAAGGAAAAAGAAATAACGCAATGGACAACTGGTGAAGCTGTGAA 910
DB 1480 AATTATAATGAAAAAGAAACTGGTGTCCACACACACAACTGGTGAAGTGTGAA 1539

QY 911 CTCAGGTGTGCACATTTATCAGGAACACACCCCAAAACCAAGTGAAGTGAAGTGAAGTGA 970
DB 1540 ATTAGGTGTGCACATTTATCTAGAAACACCCCAAAACCAAGTGAAGTGAAGTGAAGTGA 1599

QY 971 GAAGCCGTTTGTGATGTTAATTA 993
DB 1600 GAAGCTGTGTTGATGTTAATTA 1622

RESULT 6
US-09-427-048A-10
```

Sequence 10, Application US/09427048A  
Patent No. 6203975  
GENERAL INFORMATION:  
APPLICANT: Trustees of the University of Pennsylvania  
Wilson, James M.  
Fisher, Krishna J.  
Chen, Shu-Jen  
Weitzman, Matthew  
TITLE OF INVENTION: Improved Adenovirus Virus and  
Methods of Use Thereof  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Howson and Howson  
STREET: Spring House Corporate Cntr, P O Box 457  
CITY: Spring House  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19477

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/427,048A  
FILING DATE: 21-Oct-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/836,022  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: GNPVN.008PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-540-9200  
TELEFAX: 215-540-5918  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19307 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: cDNA  
SEQUENCE DESCRIPTION: SEQ ID NO: 10:  
US-09-427-048A-10

Query Match 10.2%; Score 101.4; DB 3; Length 19307;  
Best Local Similarity 81.8%; Pred. No. 1e-15;  
Matches 117; Conservative 0; Mismatches 26; Indels 0; Gaps 0;  
QY 851 AAGAAAGAAATTAATAAGGAAAGAAATACGCAATGGACAAGTGGTGAAGCTGTGAA 910  
Db 1480 AATTATAATGAAAGAAAGAACTGGTGCCACACACACAGCAAGTGGTGAAGTTGTGAA 1539  
QY 911 CTCAGGTGGCAATATACGAAACACCCCAAAAGGAGGTAGGAATAGCATGA 970  
Db 1540 ATTAGTGGCAATATACGAAACACCCCAAAAGGAGGTAGGAATAGCATGA 1599  
QY 971 GAAGCGGTGTTGATGTTAATTA 993  
Db 1600 GAAGCTGTGTTGATGTTAATTA 1622

RESULT 7  
US-08-998-416-186/c  
Sequence 186, Application US/08998416  
Patent No. 6239264  
GENERAL INFORMATION:  
APPLICANT: Philippsen, Peter  
APPLICANT: Pohlmann, Rainer  
APPLICANT: Steiner, Sabine  
APPLICANT: Mohr, Christine

APPLICANT: Wendland, Jergen  
APPLICANT: Knechtle, Philipp  
APPLICANT: Reibischung, Corinne  
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPHII  
TITLE OF INVENTION: AND USES THEREOF  
NUMBER OF SEQUENCES: 1152  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6239264artis Corporation  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: No. 6239264th Carolina  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/998,416  
FILING DATE: 24-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: CH 0016/97  
FILING DATE: 31-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 186:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 615 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: PAG1074RP  
US-08-998-416-186

Query Match 4.5%; Score 45; DB 3; Length 615;  
Best Local Similarity 48.0%; Pred. No. 0.037;  
Matches 129; Conservative 0; Mismatches 140; Indels 0; Gaps 0;  
QY 583 TATAATATTCATACAAAGAGGATTAGACAGTAGAGTTTACAAAGAAATAATCTATA 642  
Db 576 TAAATATTAATTAATAATTAATTTAAATAATAATAATAATAATAATAATAATAAT 517  
QY 643 TTTTGTGAGGAGTGGTATTATTAATCTAGTATTTTCAGTAGTTTCTAGTCTGTTATTG 702  
Db 516 TTATTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 457  
QY 703 TTTTGTTAACAATGGCAGGTTTACACGCTATGCAATTTGTACAAAAAGTTTATAAGAAA 762  
Db 456 ATCTTATAAAAGTTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 397  
QY 763 ACTACATGTAAATCTTGATGCTAAATAACITGCCATCTCTTTATATGGAACGCAATTT 822  
Db 396 ATAAAAATAATAATTTTACATATTTTAAATAATAATAATAATAATAATAATAATAAT 337  
QY 823 GGGTGTGTTAAAAATTTATAACACAGTTATA 851  
Db 336 ATTTTAATAACAATTAATAATAATAATAATA 308

RESULT 8  
US-10-204-708-31  
Sequence 31, Application US/10204708  
Patent No. 667731  
GENERAL INFORMATION:

```

; APPLICANT: OLEK, Alexander
; APPLICANT: PIERBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 31
; LENGTH: 8093
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-31

Query Match
Best Local Similarity 4.4%; Score 44; DB 4; Length 8093;
Matches 149; Conservative 0; Mismatches 155; Indels 2; Gaps 1;

QY 579 TTTTATATATTCATCAACAAGAGGATTAGACAGTAAGAGTTTACAAGATAAATC 638
Db 336 TTTTAAATATATTTATTTATAGGATTAGTAATATATATTTTAAAAA 395

QY 639 TATATTTTGTGAAGGTAGTGGTATATATCTAGATTTTCAAGTCTGTT 698
Db 396 AAGATTGTATGTATAGTAATATTTATATAGAAATATATAGAAATATATTTT 455

QY 699 ATTGTTTGTAAACATGGCAGGTTTACACGCTCTATGCAATTTGACAAAAAGTTATA 758
Db 456 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTATAG 515

QY 759 GAAAC--TACATGTAATCTGTAGTCTAAATCTGCGCATTTCTTTATATGGAACG 816
Db 516 TTAAGGGTAAATTTAAATTTTGGTTTATATTTTATTTTATTTTATTTTATATGATG 575

QY 817 CATTTTGGTGTGTTTAAATTTTAAACAGTTTAAAGAAAGAAATTTAAAGGAAAGA 876
Db 576 TAAATTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTATAGA 635

QY 877 AAATAA 882
Db 636 TTTTAA 641

RESULT 9
US-09-621-976-2813/c
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 2813
; LENGTH: 832
; TYPE: DNA

```

```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 235..399
US-09-621-976-2813

Query Match
Best Local Similarity 4.4%; Score 43.6; DB 4; Length 832;
Matches 67; Conservative 165; Mismatches 184; Indels 2; Gaps 1;

QY 576 TGTGTTTATAATATTCATACAAAGAGGATTAGACAGTAAGAGTTTACAAGATAA 635
Db 433 TGACTTTGACTATGAACATACTCCAGATTGCTAAAGTACAAGAATTGTACCATATAA 374

QY 636 ATCTATATTTTGTGAAGGTAGTGGTATTATCTAGTATTTTCAAGTCTTCAAGTCT 695
Db 373 TTAATATTTTGTGTTTWWKTTWTTTWTWMMKKARRRYYWKKSTYACASRYKKYTW 314

QY 696 G--TTATTGTTTGTAAACATGGCAGGTTTACACGCTCTATGCAATTTGACAAAAAGT 753
Db 313 GWWYMKEMSTRWVCYKCKCHYGRRCAYTWMGRWWSYAWGKWSMSMCT 254

QY 754 TATAAGAAAACTACATGTAATCTTTATAGTCTAAATCTTCCATTTCTTTATATGA 813
Db 253 RMYKKGSTYTMKCTCATWYWKYKRMWSKTCWSGSGGYMTSYTSTRSYMYWAS 194

QY 814 ACGCATTTTGGTGTGTTTAAAAATTTATAACAGTTTATAAGAAAGAAATTTATAAGAAA 873
Db 193 WMYTMCWNGRWSYTWYMAWKKWRYATTWRRAMWWAAATWYMWYMAWCKSSRGA 134

QY 874 AGAAATAACGCAATGGAACAGTGGTGAAGCTGTGAATCTGAGTGTGCAATTTATCAG 933
Db 133 AMYRRTMMWGYVYWKKSYYRTRCWAYAWKTRSYWYCWKWKRCMMMMMAAYG 74

QY 934 AACACCCCAACCAAGTGAAGTGAATATACATGAGACCGGTGTTTGTATTAAT 991
Db 73 KTMWRACWKTIRYWWAWMMRMWMTTMMTYIYWRAMKRRWMMKRSNSWMMWAW 16

RESULT 10
US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/557,884
; FILING DATE: 25-Apr-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,102
; FILING DATE: JUN-5-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PE186P3
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Query Match          4.3%; Score 43; DB 4; Length 1830121;
Best Local Similarity 48.7%; Pred. No. 1.9;
Matches 115; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 738 AATTGTACAAAAAGTTATAGAAAACTACATGTAAATCTTGATAGTAAATAACTTGC 797
Db 413354 AAGTGTATACCAATATATAAGAGATGGAATTAGTAAGATAGATAAATGCTGTT 413413
QY 798 CATTTCTTTATATGGAACGCAATTTTGGGTTGTTTAAAAATTTTAAACAGTTTATAAGAAA 857
Db 413414 TAAAGTTTAAATTTGATTTCTTATAGGTTGCTCAATTAATAATAAATCTTGATAGAAA 413473
QY 858 GAATTTATAAGAAAAAGAAATACGCAATGGAAGTGTGAGTGTGAATAGCACTCAGGT 917
Db 413474 AAATAATAGTTAATAATAACAACAATAAATGTAATGCTTGAATCTGCTGCGCAG 413533
QY 918 GTGCACAATTAATCAGGACACACCCCAAAACCAAGTGAAGTGAATAATAGCATGAGAA 973
Db 413534 TAATGTAATAAGCAACCAAGCAATPAAAAATCCAATGATATAATAANGAGAAATAAGGA 413589

```

```

RESULT 11
US-09-643-990A-1
; Sequence 1, Application US/09643990A
; Patent No. 6528289
; GENERAL INFORMATION:
; APPLICANT: Robert D. Fleischmann
; Mark D. Adams
; Owen White
; Hamilton O. Smith
; J. Craig Venter
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville,
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/643,990A
; FILING DATE: 23-Aug-2000
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,429
; FILING DATE: 1995-06-07
; APPLICATION NUMBER: 08/426,787
; FILING DATE: 1995-04-21
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB186P1C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-610-5790

```

```

; TELEFAX: 310-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Query Match          4.3%; Score 43; DB 4; Length 1830121;
Best Local Similarity 48.7%; Pred. No. 1.9;
Matches 115; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 738 AATTGTACAAAAAGTTATAGAAAACTACATGTAAATCTTTGATAGTAAATAACTTGC 797
Db 413354 AAGTGTATACCAATATATAAGAGATGGAATTAGTAAGATAGATAAATGCTGTT 413413
QY 798 CATTTCTTTATATGGAACGCAATTTTGGGTTGTTTAAAAATTTTAAACAGTTTATAAGAAA 857
Db 413414 TAAAGTTTAAATTTGATTTCTTATAGGTTGCTCAATTAATAATAAATCTTGATAGAAA 413473
QY 858 GAATTTATAAGAAAAAGAAATACGCAATGGAAGTGTGAGTGTGAAGCTGTGAACCTCAGGT 917
Db 413474 AAATAATAGTTAATAATAACAACAATAAATGTAATGCTTGAATCTGCTGCGCAG 413533
QY 918 GTGCACAATTAATCAGGACACACCCCAAAACCAAGTGAAGTGAATAATAGCATGAGAA 973
Db 413534 TAATGTAATAAGCAACCAAGCAATPAAAAATCCAATGATATAATAANGAGAAATAAGGA 413589

RESULT 12
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs

```

```
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: linear
; IMMEDIATE SOURCE:
; CLONE: PTZgpt-F1s
US-08-232-463-14

Query Match 4.3%; Score 42.8; DB 1; Length 7218;
Best Local Similarity 15.0%; Pred. No. 0.31;
Matches 41; Conservative 118; Mismatches 115; Indels 0; Gaps 0;

QY 701 TGTGTTTACAAATGGCAGGTTTACACGCTATGCAATGTACAAAAAGTTATAGA 760
Db 1518 TATTGAGTTTCAAAACCGCATGAGGCATCACTGTAATTAACCTATCTGCAAGTAGT 1459

QY 761 AAACCTACATGTAATCTTGATAGCTAAATAACCTTGCCATTTCTTTATATGGAACGCATT 820
Db 1458 TAAAGAGATAGAGAAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1399

QY 821 TTGGGTTGTTAAATTTATAACAGTTATTAAGAAAGAAATTAAGGAAAAAGAAAT 880
Db 1398 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1339

QY 881 AACGCAATGGCAAGTGGTGAAGCTGTCACTCAGTGTGCACAAATTATCAGGAACACC 940
Db 1338 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1279

QY 941 CAAACCAAGTACGAGTACAAATAGCATGAGAAG 974
Db 1278 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1245

RESULT 13
US-08-487-826B-13
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knodbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israel, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELEPHONE: (619) 235-9550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19124 base pairs
; TYPE: nucleic acid
```

```
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-487-826B-13

Query Match 4.3%; Score 42.4; DB 2; Length 19124;
Best Local Similarity 50.8%; Pred. No. 0.54;
Matches 127; Conservative 0; Mismatches 121; Indels 2; Gaps 1;

QY 640 ATATTTTTCGAGGGTAGTGGTATTATATCTAGATTTCAGTAGTTCTTAAGTCTGTTA 699
Db 4746 ATATATATTCATATGCTAGTCAATGAATATAAATATGGTATATTATATTA 4805

QY 700 TTGTTTGTAAATGCGAGGTTTACACGCTATGCAATTTGTACAAAAAGTTTAAAG 759
Db 4806 TTGTATATTAATAAGTAACACAGACATTATATATATATATTAATAATAGAATAAT 4865

QY 760 AAAACTACATGTAAATCTTGATAGCTA--AATACTTGCCTATTTCTTTATATGGAACGC 817
Db 4866 ATATTTTATGTTATATATATTTAGTTTATAAGGGGAAATTCATATATTATGAA 4925

QY 818 ATTTTGGGTTGTTTAAAAATTTATACAGTTATAAAGAAAGAAATTTAAAGGAAAAAGAA 877
Db 4926 AATTTTGTATATGATATAGTTATAAGTTAAAAAACAAGAAACAAAAATGGA 4985

QY 878 AATAACGCCAA 887
Db 4986 AAGCATAAAA 4995

RESULT 14
US-09-091-432-3
; Sequence 3, Application US/09091432
; Patent No. 5981837
; GENERAL INFORMATION:
; APPLICANT: Chapple, Clint
; TITLE OF INVENTION: A Method For Regulation Of Plant Lignin Composition
; FILE REFERENCE: 7024-325
; CURRENT APPLICATION NUMBER: US/09/091,432
; CURRENT FILING DATE: 1998-06-18
; EARLIER APPLICATION NUMBER: PCT/US96/20094
; EARLIER FILING DATE: 1996-12-19
; EARLIER APPLICATION NUMBER: US 60/009,119
; EARLIER FILING DATE: 1995-12-22
; EARLIER APPLICATION NUMBER: US 60/013,388
; EARLIER FILING DATE: 1996-03-14
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Microsoft Word 2.0C
; SEQ ID NO 3
; LENGTH: 5156
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-091-432-3

Query Match 4.2%; Score 42; DB 2; Length 5156;
Best Local Similarity 48.7%; Pred. No. 0.44;
Matches 114; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

QY 559 AACCTCTGATTCGCGCATGGTTTATAATATTATATATTCATACAAAGAGAGATTAGACAGTAA 618
Db 1156 AAGTCTTGTGAACGCACTGTATCTAAATTTGTGAGTAACGACAAATAATTTAAATTTAA 1215

QY 619 GAGTTTCAAGAAATAAATCTTATATTTTGTGAAGGGTAGTGGTATTTACTAGTATTT 678
Db 1216 GGACCTTCAAGTATTTAGTATCTCTGTCTAAGATGCACAGGTATTCAGTAATAGTAAT 1275

QY 679 CAGTAGTTTCTAAGTCTGTTATTTGTTTAAATGCGAGGTTTACACGCTCTATGCA 738
Db 1276 AATAATATCTGTATATTAATTAATCTAATAGTAACCTTGTCTTAACCTTAATGAG 1335

QY 739 ATTGTACAAAAAGTTTATAAGAAACTACATGTAATAAATCTTTGATAGCTAATAA 792
```

Db 1336 CATAAATCCAAAGCAGAAAATCTAAACCTAACTGAAAAAGTCATTACGAAAAA 1389

RESULT 15  
US-09-387-663-3  
; Sequence 3, Application US/09387663  
; Patent No. 6489538  
; GENERAL INFORMATION:  
; APPLICANT: Chapple, Clint  
; TITLE OF INVENTION: A Method For Regulation Of Plant Lignin Composition  
; FILE REFERENCE: 7024-325  
; CURRENT APPLICATION NUMBER: US/09/387,663  
; PRIOR FILING DATE: 1999-08-31  
; PRIOR APPLICATION NUMBER: US 09/091,432  
; PRIOR FILING DATE: 1998-06-18  
; PRIOR APPLICATION NUMBER: PCT/US96/20094  
; PRIOR FILING DATE: 1996-12-19  
; PRIOR APPLICATION NUMBER: US 60/009,119  
; PRIOR FILING DATE: 1995-12-22  
; PRIOR APPLICATION NUMBER: US 60/013,388  
; PRIOR FILING DATE: 1996-03-14  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: Microsoft Word 2.0C  
; SEQ ID NO 3  
; LENGTH: 5156  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-09-387-663-3

Query Match 4.2%; Score 42; DB 4; Length 5156;  
Best Local Similarity 48.7%; Pred. No. 0.44;  
Matches 114; Conservative 0; Mismatches 120; Indels 0; Gaps 0;  
QY 559 AACTCTGATCCCGCATGTTTATAATATTCATACAAAGAGGATTAGACAGTAA 618  
Db 1156 AAGTCTTGTTACGCACTGTATCTAAATGTGAGTACGCAAAAATAATTAATAATA 1215  
QY 619 GAGTTTACAGAAATAAATCTATATTTTGTGAGGGGTAGTGTATTACTGTAGATT 678  
Db 1216 GGACCTCAAGTATTATTAGTATCTCTGCTAGATGCAAGTATTCAGTAATAGTAA 1275  
QY 679 CAGTAGTTTCTAAGTCTGTTTATTGTTTGTAAATGGCAGTTTACCGTCTATGCA 738  
Db 1276 AATAATTAATCTGTATAATTAATCTAATTAGTAAACCTTGCTAAACCTAAATGAG 1335  
QY 739 ATTGTACAAAAGTTATAGAAAACACTACATGTAAATCTTGATAGCTAAATAA 792  
Db 1336 CATAAATCCAAAGCAGAAAATCTAAACCTAACTGAAAAAGTCATTACGAAAAA 1389

Search completed: April 6, 2004, 14:14:25  
Job time : 118.764 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 03:49:50 ; Search time 3674.56 Seconds  
(without alignments)  
8094.223 Million cell updates/sec

Title: US-09-966-264D-2  
Perfect score: 996  
Sequence: 1 gtggttgattgattagtaaa.....gtgttgattgatttaatt 996

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_htc:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_hic:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: em\_gss\_hum:\*
- 18: em\_gss\_inv:\*
- 19: em\_gss\_pln:\*
- 20: em\_gss\_vit:\*
- 21: em\_gss\_fun:\*
- 22: em\_gss\_mam:\*
- 23: em\_gss\_mus:\*
- 24: em\_gss\_pro:\*
- 25: em\_gss\_rod:\*
- 26: em\_gss\_phg:\*
- 27: em\_gss\_vrl:\*
- 28: gb\_gss1:\*
- 29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	459.2	46.1	525	28	AQ679243
2	443.2	44.5	472	28	AQ583263
3	410.8	41.2	844	11	BG706268
4	375.8	37.7	4437	11	AK036936

5	370.4	37.2	439	14	CB750007
6	354	35.5	924	13	BQ27942
7	342.8	34.4	629	9	AL712236
8	342.8	34.4	629	13	BX509925
9	296	29.7	633	13	BX337294
10	291.2	29.2	531	12	B1281078
11	286.2	28.7	854	13	BU226117
12	282.8	28.4	739	13	BU368789
13	282.6	28.4	500	9	AI008807
14	281.2	28.2	576	12	BG795230
15	278.6	28.0	421	13	BY635064
16	278.6	28.0	421	13	BY352875
17	271.6	27.3	308	14	N75050
18	270.8	27.2	838	13	BX354243
19	269	27.0	440	10	B8819684
20	266.2	26.7	687	14	W05777
21	264.2	26.5	929	13	BU274687
22	263.2	26.4	778	13	BU309070
23	260.2	26.1	492	14	CB725360
24	245.4	24.6	999	13	EX413180
25	245.4	24.6	648	13	BU402013
26	244.8	24.6	739	9	AU138648
27	237.2	23.8	826	13	BU104673
28	233.8	23.5	1201	13	BX463891
29	231.2	23.2	859	13	BU300671
30	229.2	23.0	928	13	BU107395
31	221.2	22.2	927	13	BU133194
32	217.8	21.9	888	14	CF553038
33	215	21.6	277	10	AW861957
34	213.8	21.5	284	10	AW604756
35	213.8	21.5	284	10	AW858457
36	208	20.9	234	14	N86611
37	202.8	20.4	619	10	AW771158
38	202.2	20.3	271	10	AW858578
39	195.6	19.6	886	13	EX768461
40	195.4	19.6	851	13	BX725372
41	194.8	19.6	628	13	BX693509
42	193.8	19.5	677	13	BX855319
43	192.8	19.4	618	9	AL793265
44	191	19.2	537	12	BU090827
45	186.6	18.7	351	13	BY673194

## ALIGNMENTS

RESULT 1	AQ679243	525 bp	DNA	linear	GSS 25-JUN-1999
LOCUS	HS_5417	Al B06	T7A	RPCI-11	Human Male BAC Library Homo sapiens
DEFINITION	genomic clone Plate=993 Col=11 Row=C, genomic survey sequence.				
ACCESSION	AQ679243				
VERSION	AQ679243.1	GI:5228047			
KEYWORDS	GSS.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	1 (bases 1 to 525)				
AUTHORS	Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.				
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	96	(17)	9739-9744	(1999)
MEDLINE	99380589				
PUBMED	10449764				
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887				

CB750007 AMGNNUC:S  
BQ27942 AGENCOURT  
AL712236 DKFZp686J  
BX509925 DKFZp686J  
BU337294 603515757  
B1281078 UI-R-DDO-  
BU226117 603947264  
BU368789 603596716  
AI008807 EST03258  
BG795230 UTSW SM30  
BY635064 BY635064  
BY352875 603525781  
N75050 2489607.81  
BU52423 603530318  
B8819684 B8819684  
W05777 2489607.r1  
BU274687 603531284  
BU309070 603538118  
CB725360 AMGNNUC:N  
EX413180 EX413180  
BU402013 604138944  
AU138648 AU138648  
BU104673 603004172  
BX463891 BX463891  
BU300671 603734536  
BU107395 603110749  
BU133194 603135045  
CF553038 AGENCOURT  
AW861957 CM3-CT034  
AW604756 CM3-CT034  
AW858457 CM3-CT034  
N86611 J9202F Huma  
AW771158 hn53c06.X  
AW858578 CM3-CT034  
EX768461 EX768461  
BX725372 BX725372  
BX693509 BX693509  
BX855319 BX855319  
AL793265 AL793265  
BU090827 BU090827  
BY673194 BY673194

Email: jwallace@u.washington.edu  
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm) or from Research Genetics (info@resgen.com). BAC end Web Server: http://www.hnsc.washington.edu  
 Plate: 993 row: C column: 11  
 Seq primer: T7  
 Class: BAC ends  
 High quality sequence stop: 525.

FEATURES  
 source  
 1. 525  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /clone="Plate=993 Col=11 Row=C"  
 /sex="male"  
 /clone\_lib="RPCI-11 Human Male BAC Library"  
 /note="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI;  
 Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"

ORIGIN  
 Query Match 46.1%; Score 459.2; DB 28; Length 525;  
 Best Local Similarity 95.4%; Pred. No. 3.4e-72; Indels 1; Gaps 1;  
 Matches 481; Conservative 0; Mismatches 22;  
 QY 137 TCTATGGAATCTTTTGATATATATTTACATGGGAACCTGAATGTAGCTTGACATTTT 196  
 DB 23 TCTCTCGGACTGCCNTGATATATATNTACATTTGGGAACCTGAATGTAGCTTGACATTTT 82  
 QY 197 CCATGTAACACCACTAGCCGATCCACATTTAGCTATCTACTATCAACACACCTGTAAT 256  
 DB 83 CCATGTTAACACCACTTTGTTGATCCACATTTAGCTATCTACTATCAACACACCTGTAAT 142  
 QY 257 GCGTTTCATTATPAGGCTTTGCTTCTCTCGAACTGGTGAAGAAATCAAACTTTGTTGT 316  
 DB 143 GCGTTTCATTATPAGGCTTTGCTTCTCTCGAACTGGTGAAGAAATCAAACTTTGTTGT 202  
 QY 317 GTACACCTCGATGAGCTTCTGTGTTCTTCCACGAAATGGGGAATGATTTCCAA 376  
 DB 203 GTACACCTCGATGAGCTTCTGTGTTCTTCCACGAAATGGGGAATGATTTCCAA 262  
 QY 377 ATGGCAAGAACAGAGTGATCTATCTATCTGACCTTTTGTAAAGTCTGTTCTTTT 436  
 DB 263 ATGGCAAGAACAGAGTGATCTATCTATCTGACCTTTTGTAAAGTCTGTTCTTTT 321  
 QY 437 CTCTTTGTTTCCAGACACAAATGAGGAAGTCTTTTCCATGCGAGATGATTTGGCA 496  
 DB 322 CTCTTTGTTTCCAGACACAAATGAGGAAGTCTTTTCCATGCGAGATGATTTGGCA 381  
 QY 497 GAGCGATGAGTCCCTTAGTATCAGTCATGACAGATGAAGAGGACAGAAATGTTT 556  
 DB 382 GAGCGATGAGTCCCTTAGTATCAGTCATGACAGATGAAGAGGACAGAAATGTTT 441  
 QY 557 ACAATCTCGATCCCGATGTTTATATATATATATATATATATATATATATATATAT 616  
 DB 442 ACAATCTCGATCCCGATGTTTATATATATATATATATATATATATATATATATAT 501  
 QY 617 AAGAGTTTACAGAAATTAATCTA 640  
 DB 502 TAGAGTTTACAGAAATTAATCTA 525

RESULT 2  
 AQ583263  
 LOCUS  
 DEFINITION  
 RPCI-11-451G14, TV RPCI-11 Homo sapiens genomic clone  
 RPCI-11-451G14, genomic survey sequence.  
 AQ583263  
 ACCESSION

VERSION  
 AQ583263.1 GI:5010373  
 GSS.  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 472)  
 Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and  
 Venter,J.C.  
 Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready  
 Map Building  
 Unpublished (1997)  
 Contact: Shaying Zhao, William Nierman, Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: hbeetig@ig.org  
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/hungen/bac\_end\_search/bac\_end\_search.html.  
 Seq primer: T7  
 Class: BAC ends.  
 Location/Qualifiers  
 1..472  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="GDB:7672957"  
 /db\_xref="taxon:9606"  
 /clone="RPCI-11-451G14"  
 /sex="Male"  
 /cell\_type="Lymphocytes"  
 /clone\_lib="RPCI-11"  
 /note="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI;  
 RPCI11 Human Male BAC Library"

ORIGIN  
 Query Match 44.5%; Score 443.2; DB 28; Length 472;  
 Best Local Similarity 96.2%; Pred. No. 2.5e-69; Indels 0; Gaps 0;  
 Matches 454; Conservative 0; Mismatches 18;  
 QY 172 AACCTGAATGTAGCTTGACATTTTCCATGTAACACCACTAGCTAGCTGATCCAACTTAAG 231  
 DB 1 AACCTGAATGTAGCTAGACATTTTCCATGTTAACACCACTAGCTAGCTGATCCAACTTAAG 60  
 QY 232 CTGATCTAACAAACACCGTGAATGGCTTCAATTAAGGCTTTGCTTCTTCTTCTTCTTCTTCTT 291  
 DB 61 CTGATCTAACAAACACCGTGAATGGCTTCAATTAAGGCTTTGCTTCTTCTTCTTCTTCTTCTT 120  
 QY 292 CTGGTGAAGAAATCAAAACCTTTGTTGTACACCTCGATGACAGCTTCTGTGTTCTTCTTCTT 351  
 DB 121 CTGGTGAAGAAATCAAAACCTTTGTTGTACACCTCGATGACAGCTTCTGTGTTCTTCTTCTT 180  
 QY 352 CCAGAAATGGGAATGATTTCCCAATGCAAGAAACAGAGTGATGCTATCTATCTGCA 411  
 DB 181 CCAGAAATGGGAATGATTTCCCAATGCAAGAAACAGAGTGATGCTATCTATCTGCA 240  
 QY 412 CCTTTTGTAAAGTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 471  
 DB 241 CCTTTTGTAAAGTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 300  
 QY 472 TTCACATGCGAGATGATTTGGCAGAGGATGGAGTCTCTTAGTATCAGTATCAGATGACAGAT 531  
 DB 301 TTCACATGCGATGATTTGGCAGAGGATGGAGTCTCTTAGTATCAGTATCAGTATGACAGAT 360  
 QY 532 GAAGAAGGACAGAAATTAATGTTTCAACTCTGCTGATTCCTGATTCCTGATTCCTGATTCCTGAT 591  
 DB 361 GAAGAAGGACAGAAATTAATGTTTCAACTCTGCTGATTCCTGATTCCTGATTCCTGATTCCTGAT 420

QY 592 CATACAAAGAGGATAGACAGTACAGTATTTTACAGAAATAAATCTATAT 643  
 Db 421 CATACAAAGAGGATAGACAGTATTTTACAGAAATAAATCTATAT 472

RESULT 3  
 BG706268 602669567F1 NIH\_MGC\_96 Homo sapiens cDNA clone IMAGE:4792313 5',  
 LOCUS mRNA sequence.  
 DEFINITION  
 ACCESSION BG706268  
 VERSION BG706268.1 GI:13981445  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 TITLE 1 (bases 1 to 844)  
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
 COMMENT Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
 Toshlyuki and Piero Carninci (RIKEN)  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: L1AM10670 row: e column: 18  
 High quality sequence stop: 558.  
 Location/Qualifiers

1. 844  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4792313"  
 /tissue\_type="hypothalamus"  
 /lab\_host="DH10B"  
 /clone\_lib="NIH\_MGC\_96"  
 /notes="Organ: brain; Vector: pBluescriptR (modified  
 pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI  
 (gtcag); Oligo-dT primed using primer  
 5'-TTTTTTTTTTTTTTT-3', size-selected for average  
 insert size 2.3 kb and normalized to ROT 5. This is a  
 primary library enriched for full-length clones and  
 constructed using the Cap-trapper method (Carninci, in  
 preparation). Library constructed by M. Brownstein  
 (NIH/NHGRI, National Institutes of Health). Note: this is  
 a NIH\_MGC Library."  
 ORIGIN  
 Query Match 41.2%; Score 410.8; DB 12; Length 844;  
 Best Local Similarity 98.3%; Pred. No. 1.2e-63;  
 Matches 415; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 450 AGGACACAATGTAGGAAGTCTTTTCCCATGGCAGATCAATTTGGCAGAGCGATGAGTC 509  
 Db 88 AGGACACAATGTAGGAAGTCTTTTCCCATGGCAGATCAATTTGGCAGAGCGATGAGTC 147  
 QY 510 CTTAGTATCAGTCATGACAGATGAAGAGGAGGAGAGTAATGTTTACAACTCTTGATT 569  
 Db 148 CTTAGTATCAGTCATGACAGATGAAGAGGAGGAGAGTAATGTTTACAACTCTTGATT 207  
 QY 570 CCCGATGTTTTTATATATTCATACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629  
 Db 208 CCCGATGTTTTTATATATTCATACAAAGAGGATTAGACAGTAAGAGTTTACAAG 267  
 QY 630 AAATAAATCTATATTTTGTGAAGGGTAGTGGTATTAATCTAGATTTTCAAGTATTTCT 689  
 Db 268 AAATAAATCTATATTTTGTGAAGGGTAGTGGTATTAATCTAGATTTTCAAGTATTTCT 327

QY 690 AAGTCTGTTATTGTTTGTAACTAATGCGAGTTTACAGTCTATGCAATTTGACAAA 749  
 Db 328 AAGTCTGTTATTGTTTGTAACTAATGCGAGTTTACAGTCTATGCAATTTGACAAA 387  
 QY 750 AAGTTATTAAGAAACTACATCTAAATCTTCATAGCTAAATAAATGTCATTTCTTATA 809  
 Db 388 AAGTTATTAAGAAACTACATCTAAATCTTCATAGCTAAATAAATGTCATTTCTTATA 447  
 QY 810 TGGACGCAATTTGGGTTGTTTAAATAATTAACAGTTATTAAGAAAGAAATTAAGG 869  
 Db 448 TGGACGCAATTTGGGTTGTTTAAATAATTAACAGTTATTAAGAAAGAAATTAAGG 507  
 QY 870 AA 871  
 Db 508 AA 509

RESULT 4  
 AK036936 4437 bp mRNA linear HTC 19-SEP-2003  
 LOCUS Mus musculus adult female vagina cDNA, RIKEN full-length enriched  
 DEFINITION library, clone:9930028B14 product:dystrophin, muscular dystrophy,  
 full insert sequence.

ACCESSION AK036936  
 VERSION AK036936.1 GI:26085504  
 KEYWORDS HTC; CAP trapper.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus

REFERENCE  
 AUTHORS Carninci, P. and Hayashizaki, Y.  
 TITLE High-efficiency full-length cDNA cloning  
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
 MEDLINE 99279253  
 PUBMED 10349636

REFERENCE  
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes  
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
 MEDLINE 20499374  
 PUBMED 11042159

REFERENCE  
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
 Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,  
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, J.,  
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,  
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 TITLE RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer  
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
 MEDLINE 20530913  
 PUBMED 11076861

REFERENCE  
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the  
 FANTOM Consortium.  
 TITLE Functional annotation of a full-length mouse cDNA collection  
 JOURNAL Nature 409, 685-690 (2001)  
 REFERENCE

REFERENCE  
 AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research  
 Group Phase I & II Team.  
 TITLE Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 JOURNAL Nature 420, 563-573 (2002)  
 REFERENCE

REFERENCE  
 AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,  
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,  
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,

Hori, F., Imotani, K., Iehii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ono, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.

Direct Submission  
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216]

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.  
URL: http://genome.gsc.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/  
Location/Qualifiers  
1. .4437  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="PANTOM\_DB:9930028B14"  
/db\_xref="MGI:2401399"  
/db\_xref="taxon:10090"  
/clone="9930028B14"  
/sex="female"  
/tissue\_type="vagina"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"  
29. .1988  
/note="dystrophin, muscular dystrophy (MGD|MGI:94909, GB|NM\_007868, evidence: BLASTN, 100%, match=99); putative"  
polyA\_signal  
4417. .4422  
/note="putative"  
polyA\_site  
4437  
/note="putative"

misc\_feature  
29. .1988  
/note="dystrophin, muscular dystrophy (MGD|MGI:94909, GB|NM\_007868, evidence: BLASTN, 100%, match=99); putative"

ORIGIN  
Query Match 37.7%; Score 375.8; DB 11; Length 4437;  
Best Local Similarity 96.2%; Pred. No. 1e-57;  
Matches 407; Conservative 0; Mismatches 12; Indels 4; Gaps 2;  
QY 450 AGGACACATGTAGGAAGTCCTTTCCACATGCGCAGATGTTGGCAGAGCGATGGAGTC 509  
DB 1869 AGGACACATGTAGGAAGTCCTTTCCACATGCGCAGATGTTGGCAGAGCGATGGAGTC 1928  
QY 510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGCAATAATGTTTACAACCTCTGATT 569  
DB 1929 CTTAGTTCAGTCATGACAGATGAAGAAGGAGCAGCAATAATGTTTACAACCTCTGATT 1988  
QY 570 CCGCATGTTTATATATATTCATACAAAGAGGATTAGACAGTAAGATTTACAAAG 629  
DB 1989 CCGCATGTTTATATATATTCATACAAAGAGGATTAGACAGTAAGATTTACAAAG 2048  
QY 630 AAAT-AAAATCTATATTTTGTGAAGGAGTGTGTTATATCTGTAGATTTCAGTAGTTTC 688  
DB 2049 AAATAAATCTATATTTTGTGAAGGAGTGTGTTATCTGTAGATTTCAGTAGTTTC 2108  
QY 689 TAAGTCTGTATGTTTGTGTTAAACAGAGGTTTACAGTCTATGCAATTTGACAAA 748  
DB 2109 TAAGTCTGTATGTTTGTGTTAAACAGAGGTTTACAGTCTATGCAATTTGACAAA 2168  
QY 749 AAAGTATAAGAAACTACATCTAAATCTTGTAGCTAAATAAATCTGCCATTTCTTTAT 808

DB 2169 AAAGTATAAGAAACTACATCTAAATCTTGTAGCTAAATAAATCTGCCATTTCTTTAT 2225  
QY 809 ATGGAACGCAATTTGGTGTGTTTAAAAATTTATAACAGTTTATAAGAAAGATTTAAAG 868  
DB 2226 ATGGAACGCAATTTGGTGTGTTTAAAAATTTATAACAGTTTATAAGAAAGATTTAAAG 2285  
QY 869 GAA 871  
DB 2286 AAA 2288

RESULT 5  
CB750007  
LOCUS  
DEFINITION  
AMGNNUC:SRPB2-00166-B9-A.srp2 (10220) Rattus norvegicus cDNA clone  
srp2-00166-b9 5', mRNA sequence.  
CB750007  
CB750007.1 GI:29817309  
VERSION  
ACCESSION  
Rattus norvegicus (Norway rat)  
KEYWORDS  
SOURCE  
ORGANISM  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 439)  
Angen EST Program.  
Angen Rat EST Program  
Unpublished (2003)  
Contact: Dan Fitzpatrick  
Angen, Inc  
One Angen Center Drive, Thousand Oaks, CA 91320-1799, USA  
Tel: 805 447-4881  
Plate: 00166 row: b column: 9.  
Location/Qualifiers  
1. .439  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10116"  
/clone="srp2-00166-b9"  
/tissue\_type="prostate tissue"  
/clone\_lib="srp2 (10220)"  
/note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; rat  
prostate normalized double selected poly(A+) mRNA size  
fraction > 1 kb"

ORIGIN  
Query Match 37.2%; Score 370.4; DB 14; Length 439;  
Best Local Similarity 95.4%; Pred. No. 2.4e-56;  
Matches 393; Conservative 0; Mismatches 16; Indels 3; Gaps 1;  
QY 460 GTAGGAAGTCCTTTCCACATGCGCAGATGTTGGCAGAGCGATGGAGTCCTTAGTATCA 519  
DB 12 GTAGGAAGTCCTTTCCACATGCGCAGATGTTGGCAGAGCGATGGAGTCCTTAGTATCA 71  
QY 520 GTCATGACAGATGAAGAAGGAGCAGCAATAATGTTTACAACCTCTGATCCCGCATGT 579  
DB 72 GTCATGATAGATGAAGAAGGAGCAGCAATAATGTTTACAACCTCTGATCCCGCATGT 131  
QY 580 TTTTATAATTCATACAAAGAGGATTAGACAGTAAGAGTTTACAGAAATAAATCT 639  
DB 132 TTTTATAATTCATACAAAGAGGATTAGACAGTAAGAGTTTACAGAAATAAATCT 191  
QY 640 ATATTTTGTGAAGGAGTGTGTTATATCTGTAGATTTCAGTAGTTTCAGTAGTTTCAGT 699  
DB 192 ATATTTTGTGAAGGAGTGTGTTATCTGTAGATTTCAGTAGTTTCAGTAGTTTCAGT 251  
QY 700 TTTTGTGTAAAGTGTGTTTACAGCTCTGCAATGTTGACAAAAGATTTAAG 759  
DB 252 TTTTGTGTAAAGTGTGTTTACAGCTCTGCAATGTTGACAAAAGATTTAAG 311  
QY 760 AAAACTACATGTAAAATCTTGTAGCTAAATAAATCTGCCATTTCTTTATATGGAACGAT 819

```

Db      312 AAA---ACATGTAATACTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCAT 368
Qy      920 TTGGGGTGTGTTAAATAATTTATAACAGTTTATAAGAAAGAAATTAAGGAA 871
Db      369 TTGGGGTGTGTTAAATAATTTATAACAGTTTATAAGAAAGAAATTTGAACTAAA 420

RESULT 6
BQ927942      BQ927942      924 bp      mRNA      linear      EST 20-AUG-2002
LOCUS      AGENCOURT_8858741 NCI_CGAP Mam2 Mus musculus cDNA clone
DEFINITION      IMAGE:6441872 5', mRNA sequence.
ACCESSION      BQ927942
VERSION      BQ927942.1 GI:22342973
KEYWORDS      Mus musculus (house mouse)
SOURCE      Mus musculus
ORGANISM      Mus musculus
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE      NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL      National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT      Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.,
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13965 row: m column: 09
High quality sequence start: 10
High quality sequence stop: 552.

FEATURES             Location/Qualifiers
     source             1..924
     -organism="Mus musculus"
     -mol_type="mRNA"
     -strain="FVB/N-3"
     -db_xref="taxon:10090"
     -clone="IMAGE:6441872"
     -tissue_type="tumor, biopsy sample"
     -dev_stage="5 months"
     -lab_host="DH10B"
     /clone_lib="NCI CGAP Mam2"
     /note="Organ: mammary; Vector: pCMV-SPOK76; Site_1: SalI;
     Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
     Library constructed by Life Technologies. Investigator
     providing samples: Gilbert Smith, NIH"

ORIGIN
Query Match      35.5%; Score 354; DB 13; Length 924;
Best Local Similarity 95.3%; Pred. No. 1.5e-53;
Matches 387; Conservative 0; Mismatches 15; Indels 4; Gaps 2;

Qy      467 GTCTTTTCCATGCGCAGATGTTGGCAGAGCGATGGAGTCTTAGTATCAGTCATGA 526
Db      33  GCGCTTTTCCCATGCGCAGATGTTGGCAGAGCGATGGAGTCTTAGTATCAGTCATGA 92
Qy      527 CAGATGAAGAGGACAGAAATAAATGTTTCAACTCTGATTCGCCGATCGGTTTTATA 586
Db      93  CAGATGAAGAGGACAGAAATAAATGTTTCAACTCTGATTCGCCGATCGGTTTTATA 152
Qy      587 ATATTTCATACAAAGAGGATTACACAGTAGAGTTTACAGAAAT-AAATCTATATTT 645
Db      153 ATATTTCGACAAAGAGGATTACACAGTAGAGTTTACAGAAATATAATCTATATTT 212
Qy      646 TTGTGAAGGGTAGTGGTATTATPACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATGTTT 705
Db      213 TTGTGAAGGGTAGTGGTACTATPACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATGTTT 272
Qy      706 TGTAAACATGCGCAGTTTACACGCTCTATGCAATGTTACAAAAGTTTAAGAAACT 765

```

```

Db      273 TGTAAACATGCGCAGTTTACACGCTCTATGCAATTTGTACAAAAGTTTAAAGAAA--- 329
Qy      766 ACATGTAATACTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTGGG 825
Db      330 ACATGTAATACTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTGGG 389
Qy      826 TTGTTTAAAAATTTATAACAGTTTATAAGAAAGAAATTTATAAGGAA 871
Db      390 TTGTTTAAAAATTTATAACAGTTTATAAGAAAGAAATTTATAAGTAACTAAA 435

RESULT 7
AL712236      629 bp      mRNA      linear      EST 04-SEP-2003
LOCUS      DKFZp686J0287.5', mRNA sequence.
DEFINITION      AL712236
ACCESSION      AL712236
VERSION      AL712236
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE      Pou5f1 (Pou5f1), Albert, R., Moosmayer, P., Schupp, I., Wellenreuther, R.,
JOURNAL      Wewes, H.W., Well, B., Amid, C., Osanger, A., Pobo, G., Han, M. and
COMMENT      Wiemann, S.
EST (Pou5f1), Albert, R., Moosmayer, P., Schupp, I., Wellenreuther, R.,
Unpublished (2003)
On Mar 22, 2002 this sequence version replaced gi:19695591.
Contact: MIPS
MIPS
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by DKFZ (German Cancer Research Center,
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.
No si sequence available.
This clone (DKFZp686J0287) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES             Location/Qualifiers
     source             1..629
     -organism="Homo sapiens"
     -mol_type="mRNA"
     -db_xref="taxon:9606"
     -clone="DKFZp686J0287"
     -dev_stage="adult"
     -lab_host="DH10B"
     -clone_lib="686 (synonym: hlcc3)"
     /note="Vector: pTriplex2; Site_1: SfiI; Site_2: SfiI;
     cDNA-collection"

ORIGIN
Query Match      34.4%; Score 342.8; DB 9; Length 629;
Best Local Similarity 98.0%; Pred. No. 1.7e-51;
Matches 347; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy      518 CAGTCATGACAGATGAAGAGGAGCAGATAAATGTTTACAACTCCTGATTCGCCCATG 577
Db      1  CAGTCATGACAGATGAAGAGGAGCAGATAAATGTTTACAACTCCTGATTCGCCCATG 60
Qy      578 GTTTTTTATATTCATACAAAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAT 637
Db      61  GTTTTTTATATTCATACAAAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAT 120
Qy      638 CTATATTTTGTGAAGGGTAGTGGTATTACTGTAGATTTCAGTAGTTTCTAAGTCTCT 697
Db      121 CTATATTTTGTGAAGGGTAGTGGTATTACTGTAGATTTCAGTAGTTTCTAAGTCTCT 180

```

```

QY 698 TATTGTTTGTAAATGCGAGGTTTACAGCTCTATGCAATTTGTACAAAAAGTTTATA 757
DB 181 TATTGTTTGTAAATGCGAGGTTTACAGCTCTATGCAATTTGTACAAAAAGTTTATA 240
QY 758 AGAAAACTACATGTAATCTTTGATAGTAAATTAACCTTGGCAATTTCTTTATATGGAACGC 817
DB 241 AGAAAACTACATGTAATCTTTGATAGTAAATTAACCTTGGCAATTTCTTTATATGGAACGC 300
QY 818 ATTTTGGTGTGTTTAAAAATTTTAAACAGTTTATAAAGAAATTAATTAAGGAA 871
DB 301 ATTTTGGTGTGTTTAAAAATTTTAAACAGTTTATAAAGAAATTTGTAACCTAAA 354

RESULT 8
BX509925 629 bp mRNA linear EST 04-SEP-2003
LOCUS DKFZP686J2487_r1 686 (synonym: hicc3) Homo sapiens cDNA clone
DEFINITION DKFZP686J2487 5', mRNA sequence.
ACCESSION BX509925
VERSION BX509925
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 629)
AUTHORS Poustka,A., Albert,R., Moosmayer,P., Schupp,I., Wellenreuther,R.,
Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
Wiemann,S.
TITLE EST (Poustka,A., Albert,R., Moosmayer,P., Schupp,I.,
Wellenreuther,R., et al.)
JOURNAL Unpublished (2003)
COMMENT Contact: MIPS
MIPS Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.
No s1 sequence available.
This clone (DKFZP686J2487) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES
source
1..629
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZP686J2487"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="hicc3"
/note="Vector: pTriplex2; Site_1: sf11A; Site_2: sf11B;
cDNA-collection"

ORIGIN
Query Match 34.4%; Score 342.8; DB 13; Length 629;
Best Local Similarity 98.0%; Pred. No. 1.7e-51;
Matches 347; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 518 CAGTCATGACAGATGAAGAGGAGCAGATAATGTTTACAACTCCGATTCGCGCATG 577
DB 1 CAGTCATGACAGATGAAGAGGAGCAGATAATGTTTACAACTCCGATTCGCGCATG 60
QY 578 GTTTTATATATCTATACAAAGAGGATTAGACAGTAAGAGTTTACAGAAATAAT 637
DB 61 GTTTTATATATCTATACAAAGAGGATTAGACAGTAAGAGTTTACAGAAATAAT 120
QY 638 CTATATTTTGTCAAGGTTAGTGTATATCTGTAGATTTCTAGTAGTTTCTAAGTCTGT 697

```

```

DB 121 CTATATTTTGTGAAGGTTAGTGGTATTATCTAGTAGATTTTCAGTAGTTTCTAAGTCTGT 180
QY 698 TATTGTTTGTAAATGCGAGGTTTACAGCTCTATGCAATTTGTACAAAAAGTTTATA 757
DB 181 TATTGTTTGTAAATGCGAGGTTTACAGCTCTATGCAATTTGTACAAAAAGTTTATA 240
QY 758 AGAAAACTACATGTAATCTTTGATAGTAAATTAACCTTGGCAATTTCTTTATATGGAACGC 817
DB 241 AGAAAACTACATGTAATCTTTGATAGTAAATTAACCTTGGCAATTTCTTTATATGGAACGC 300
QY 818 ATTTTGGTGTGTTTAAAAATTTTAAACAGTTTATAAAGAAATTAATTAAGGAA 871
DB 301 ATTTTGGTGTGTTTAAAAATTTTAAACAGTTTATAAAGAAATTTGTAACCTAAA 354

RESULT 9
BX337294 633 bp mRNA linear EST 28-NOV-2002
LOCUS 60351575F1 CSEQCHN66 Gallus gallus cDNA clone Chest455a15 5', mRNA
DEFINITION sequence.
ACCESSION BU337294
VERSION BU337294
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 633)
AUTHORS Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
TITLE A Comprehensive Collection of Chicken cDNAs
JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE 22335534
PUBMED 12445392
COMMENT Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
FEATURES
source
1..633
Location/Qualifiers
/organism="Gallus gallus"
/mol_type="mRNA"
/db_xref="taxon:9031"
/clone="Chest455a15"
/dev_stage="16 day embryo"
/lab_host="DH10B"
/clone_lib="CSEQCHN66"
/note="Organ: brain; Vector: pBluescript II KS(+); Site_1:
ECORI; Site_2: NotI; this normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer, using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunted, ligated to NotI adapters, digested with
EcoRI. Size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research
(1996): 791, except that a significantly longer
reannealing hybridization was used."

ORIGIN
Query Match 29.7%; Score 296; DB 13; Length 633;
Best Local Similarity 87.2%; Pred. No. 3.8e-43;
Matches 396; Conservative 0; Mismatches 45; Indels 13; Gaps 6;

```

431	QY	TTCTTTCTCTTTGGTTTTCCAGGACAAATGTAGGAAGTCTTTTCCCACTGGCAGATGATT	490
140	Db		
140	Db	TTCTTTGGCTTTCTCTTTTCCAGGCCACAATGTAGGGAGCCTTTTCCCACTGGCAGATGACT	199
491	QY	TCGGCAGAGCGATGGAGTCCTTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAGAT---	547
200	Db		
200	Db	TCGGAAGAGCGATGGAACCTTAGTGACGGTGATGACTGACGACAGGTGTTAGATAGC	259
548	QY	---AAATGTTTTACAAC--TCCTGATTCGGCATCGTTTTTATAATATTCATACACAAG	603
260	Db	AAAGATGTGTTTTACAACCTCAGGCGTCGGCATGGTTTTTATAATATTCATACTACAAAG	319
604	QY	AGGATTAGACAGTAGAGTTTACAG--AAATAAATCTATATTTTGTGAAGGCTAGTGGT	662
320	Db	AGGATTAGACTGTAGAGTTTACAGAAAAAAATCTATATTTTGTGAAGGCTAGTGGT	379
663	QY	ATTATACTGTAGATTTCAAGTAGTTT--CTAAGCTGTGTTATTGTTTTTAACAATGGCAG	721
380	Db	ACTATACTGTAGATTTCAAGTAGTTTCCTAAGTCGTGTTCTGTTTTGTTAACAATGGCAG	439
722	QY	TTTTTACACGCTATGCAATGTGTACAAAAAGTTTATAGAAAACTACATGTAAAAATCTTGA	781
440	Db	TTTTTACACGCTATGCAATGTGTACAAAAA---TTTAAAGAAAACTACATGTAAAAATCTTGA	497
782	QY	TAGCTAAATAACTGTCGCAATTCCTTTATATGAAACGCAATTTTGGGTGCTTTT--AAAAATTT	839
498	Db	TAGCTGAATAACTGTCGCAATTCCTTTATATGAAACGCAATTTTGGGTGCTTTTAAAAAATTT	557
840	QY	ATAACAGTTTATAAGAAAGAAATTAATAAGGAAAA	873
558	Db	ATAACAGTTTATAAGAAAGAAATCGTAAACTATAAAA	591

RESULT 10	531 bp	linear	EST 19-JUL-2001
B1281078/c			
B1281078/c			
LOCUS			
DEFINITION			
	B1281078	531 bp	linear
	UI-R-DD0-b2x-a-02-0-UI.s1	UI-R-DD0	Rattus norvegicus cDNA clone
	UI-R-DD0-b2x-a-02-0-UI.3'		mRNA sequence.
	UI-R-DD0-b2x-a-02-0-UI.3'		mRNA sequence.

ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
B1261078	B1281078_1	GI:14930459	Rattus norvegicus (Norway rat)	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
		EST.	Rattus norvegicus	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
			Rattus norvegicus	Rattus

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

1 (bases 1 to 531)  
Bonaldo, M.F., Lennon, G. and Soares, M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9). 791-806 (1996)

MEDLINE PUBMED	97044377 869548	Contact: Soares, MB Coordinated Laboratory for Computational Genomics University of Iowa 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA Tel: 319 335 8250 Fax: 319 335 9585
-------------------	--------------------	--

Email: bento-soares@uiowa.edu  
 Fax: 319.335.3000  
 The sequence contained an oligo-dT track that was present in the  
 oligonucleotide that was used to prime the synthesis of first  
 strand cDNA and therefore this may represent a bonafide poly A  
 tail. The sequence tag present in the cDNA between the NotI site  
 and the oligo-dT track served to verify it as a clone from the  
 non-normalized fundus library cDNA library Preparation: M.B. Soares  
 Lab Clone distribution: clones will be available through Research  
 Genetics (www.resgen.com)  
 Seq primer: M13 Forward  
 POLYA=yes.

```

FEATURES      Location/Qualifiers
source        1..531
               /organism="Rattus norvegicus"

```

```

/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-DDO-bzr-a-02-O-UI"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-DDO"
/notes"Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not 1; Site 2: Eco RI; The UI-R-DDO
library is a non-normalized library constructed from rat
funds tissue. For a detailed description of the library
from which this clone was derived, please visit our web
site at ratest.eng.uiowa.edu. The subtraction has been
previously described in (Bonaldo, Lennon and Soares,
Genome Research 6:791-806, 1996)
TAG_LIB=UI-R-DDO
TAG_TISSUE=fundus
TAG_SEQ=TTCCGG"

```

## ORIGIN

```
Query Match      29.2%; Score 291.2; DB 12; Length 531;
Best Local Similarity 95.8%; Pred. No. 2.9e-42;
Matches 299; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
```

Qy	450	AGGACCAATGTTAGGAAGTCTTTTCCACATGGCAGATGATTTGGCCAGCAGCGATGGAGTC	500
Db	312	AGGACCAATGTTAGGAAGCCTTTTCCACATGGCAGATGATTTGGCCAGCAGCGATGGAGTC	253
Qy	510	CTTAGTATCAGTCATGACAGATGAAGGAAGGACAGAAATAATGTTTTCACACTCCTGATT	569
Db	252	CTTAGTTTCAGTCATGATAGATGAAGGAAGGACAGAAATAATGTTTTCACACTCCTGACC	193
Qy	570	CCGCGATGGTTTTATTAATTCATACAAACAAGAGGATTAGACAGATAAGAGTTTACAAAG	629
Db	192	CCGCGCATGGTTTTATTAATTTCTGACACAAGAGGATTTAGACAGTAAGCGTTTACAAAG	133
Qy	630	AAATAAATCTATATTTTGTGAAGGGTAGTGGTATATATACGTAGATTTTCAGTAGTTTCT	689
Db	132	AAATAAATCTATATTTTGTGAAGGGTAGTGGTACTATCTGTAGATTTTCAGTAGTTTCT	73
Qy	690	AAGTCTGTTATGTTTTGTTAAACAATGCAGGTTTTTACAGCTCATGCAATTTGTACAAAA	749
Db	72	AAGTCTGTTATGTTTTGTTAAACAATGCAGGTTTTTACAGCTCATGCAATTTGTACAAAA	13
Qy	750	AAGTTATAAGAA 761	
Db	12	AAAAAAAAAAAAA 1	

## RESULT 11

BU226117

**LOCUS**

### DEFINITION

**ACCESSION**

VERSION

## KEYWORDS

**SOURCE**

ORGANISM

CONFIDENTIAL

## REFERENCES

## CONCLUSION

**TITLE**

JOURNAL

MEDLINE

PUBMED

COMMENT

BU226117 854 bp mRNA linear EST 26-NOV-2002  
603947264F1 CSEQCHN23 Gallus gallus cdna clone CHEST901p15 5', mRNA  
sequence.  
BU226117  
BU226117.1 GI:25462707  
EST.  
Gallus gallus (chicken)  
Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
1 (bases 1 to 854)  
Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,  
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.  
A Comprehensive Collection of Chicken cDNAs  
Curr. Biol. 12 (22), 1965-1969 (2002)

12445392  
Contact: Simon Hubbard  
Department of Biomolecular Sciences  
University of Manchester Institute of Science and Technology  
(UMIST)

PO Box 88, Manchester, M60 1QD, UK  
Tel: 01612008930  
Fax: 01612360409  
Email: Simon.Hubbard@umist.ac.uk.  
Location/Qualifiers

# FEATURES

source  
1. .854  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="White Leghorn, HiseX"  
/db\_xref="taxon:9031"  
/clone="ChEST901p15"  
/dev\_stage="22"  
/lab\_host="DH10B"  
/clone\_lib="CSEQCHN23"  
/note="Organ: heads; Vector: pBluescript II KS(+); Site\_1:  
EcoRI; Site 2: NotI; This normalized library was  
constructed from 1 million independent clones. cDNA  
synthesis was initiated using an oligo(dT) primer, using  
methylated C in the first strand synthesis reaction.  
Following this first strand reaction, double-stranded cDNA  
was blunted, ligated to NotI adapters, digested with  
EcoRI, size-selected, and cloned into the NotI and EcoRI  
compatible sites of a custom modified MCS of the  
pBluescript (KS+) vector. The library was normalized in 2  
rounds using conditions adapted from Soares et al., PNAS  
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6  
(1996): 791, except that a significantly longer  
reannealing hybridization was used."

# ORIGIN

Query Match 28.7%; Score 286.2; DB 13; Length 854;  
Best Local Similarity 86.3%; Pred. No. 1.8e-41;  
Matches 389; Conservative 0; Mismatches 48; Indels 14; Gaps 6;  
QY 437 CTCCTTTGTTTCCAGGACACATGATGAGAGCTTTTCCATGCGCAGATGATGGCA 496  
DB 5 CTCGAGGTTCCAGAGGCCACATGATGAGGAGCTTTTCCATGCGCAGATGATGGAA 64  
QY 497 GAGCGATGAGTCCCTAGTATCATGACATGATGAGAGGAGGAGCAAT-----AAA 550  
DB 65 GAGCGATGAGAACCTTAGTGACGGTGATGACGACGACAGAGGTGTAGATAGCAAGATG 124  
QY 551 TGTTTTACAC--TCCTGATCCCGATGTTTATATATATATATATATATATATATATATAT 609  
DB 125 TGTTTTACAACTTCAGCGCTCCCGCATGTTTATATATATATATATATATATATATATAT 184  
QY 610 AGACATGATGAGTTTACAG--AAATAATCTATATTTTGTGAGCGGTAGTGGTATTATA 668  
DB 185 AGACTGTAAAGTTTACAGAAACAAATCTACATTTTGTGAGGGGTAGTGGTACTATA 244  
QY 669 CTGTAGATTTCAGTAGTTT--CTAAGTCTGTTATTTGTTTGTAAACAATGGCAGGTTTAC 727  
DB 245 CTGTAGATTTCAGTAGTTTCTTAAAGTCTGTTTGTGTTTGTAAACAATGGCAGGTTTAC 304  
QY 728 AGCTCTATGCAATGTACAAAAGTTTAAAGAACTACATGTAATCTTTGATAGCTA 787  
DB 305 AGCTCTATGCAATGTACAAA--TTAAGAACTACATGTAATCTTTGATAGCTG 362  
QY 788 AATAACTTGCATTTCTTTATATGGAACGCAITTTGGGTTGTTT---AAAAATTTATAAC 844  
DB 363 AATAACTTGCATTTCTTTATATGGAACGCAITTTGGGTTGTTTAAAAAATTTATAAC 422  
QY 845 AGTTATTAAGAAAGANTTAAAGGAAAAG 875  
DB 423 AGTTATTAAGAAAGAGATTGTAACATAAAG 453

# RESULT 12

BU368789  
LOCUS 739 bp, mRNA linear EST 28-NOV-2002  
DEFINITION 603596716F1 CSEQCHN73 Gallus gallus cDNA clone ChEST56503 5', mRNA  
sequence.  
ACCESSION BU368789

# VERSION KEYWORDS SOURCE ORGANISM

BU368789.1 GI:25876790  
EST.  
Gallus gallus (chicken)  
Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.

# REFERENCE

1 (bases 1 to 739)  
Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,  
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.  
A Comprehensive Collection of Chicken cDNAs  
Curr. Biol. 12 (22), 1965-1969 (2002)

# JOURNAL MEDLINE PUBMED

22335534  
12445392

# COMMENT

Contact: Simon Hubbard  
Department of Biochemical Sciences  
University of Manchester Institute of Science and Technology  
(UMIST)  
PO Box 88, Manchester, M60 1QD, UK  
Tel: 01612008930  
Fax: 01612360409  
Email: Simon.Hubbard@umist.ac.uk.  
Location/Qualifiers

# FEATURES

source

1. .739  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="Compton Line 151"  
/db\_xref="taxon:9031"  
/clone="ChEST56503"  
/sex="Female"  
/tissue\_type="not cerebrum or cerebellum"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/clone\_lib="CSEQCHN73"  
/note="Organ: brain; Vector: pBluescript II KS(+); Site\_1:  
EcoRI; Site 2: NotI; This normalized library was  
constructed from 1 million independent clones. cDNA  
synthesis was initiated using an oligo(dT) primer, using  
methylated C in the first strand synthesis reaction.  
Following this first strand reaction, double-stranded cDNA  
was blunted, ligated to NotI adapters, digested with  
EcoRI, size-selected, and cloned into the NotI and EcoRI  
compatible sites of a custom modified MCS of the  
pBluescript (KS+) vector. The library was normalized in 2  
rounds using conditions adapted from Soares et al., PNAS  
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6  
(1996): 791, except that a significantly longer  
reannealing hybridization was used."

# ORIGIN

Query Match 28.4%; Score 282.8; DB 13; Length 739;  
Best Local Similarity 86.5%; Pred. No. 7.9e-41;  
Matches 384; Conservative 0; Mismatches 47; Indels 13; Gaps 6;  
QY 443 GTTTTCCAGGACACAATGTAGGAAGTCTTTCCATGCGCAGATGATTTGGGACAGCGA 502  
DB 256 GTTCCAGAGGCCACAATGTAGGAGCGCTTTCCATGCGCAGATGATTTGGGACAGCGA 315  
QY 503 TGGAGTCTTAGTATCAGTCTACAGATGACAGAGGACAGCAAT-----AAATGTTTT 556  
DB 316 TGGAAACCTTAGTGACGGTGTATGATGACGACAGAGGTGTAGANTAGCAAGATGTTTT 375  
QY 557 ACAAC-TCCTGATTCCTGATGCTTTTATATATATATATATATATATATATATATATAT 615  
DB 376 ACACTTCAGCGCTCCCGCATGTTTATATATATATATATATATATATATATATATAT 435  
QY 616 TAAGAGTTTACAG-AAAATAATCTATATTTTGTGAGGGGTAGTGGTATTACTGTAG 674  
DB 436 TAAGAGTTTACAGAAACAAATCTATATTTTGTGAGGGGTAGTGGTATTACTGTAG 495  
QY 675 ATTTTCAGTAGTTT--CTAAGTCTGTTATTTGTTTGTAAACAATGGCAGGTTTTACAGTCT 733  
DB 496 ATTTTCAGTAGTTTCTTAAAGTCTGTTTCTGTTTGGTTAAACAATGGCAGGTTTTACAGTCT 555



```

QY 734 ATGCAATTGTACAAAAAGTTTATAGAAAACTACATGTAATAAATCTTGATAGCTAAATAAC 793
Db 556 ATGCAATTGTACAAAA--TTATAGAAAACTACATGTAATAAATCTTGATAGCTGAATAAC 613
QY 794 TTGCATTTCTTTATATAGAACGATTTTGGTGTGTTA--AAATTTATAACAGTTATA 851
Db 614 TTGCATTTCTTTATATAGAACGATTTTGGTGTGTTA--AAATTTATAACAGTTATA 673
QY 852 AAGAAAGAAATTATAAGGAAAAAG 875
Db 674 AAGAAAGATCTAACTAAAAAG 697

RESULT 13
LOCUS AI008807/c
DEFINITION EST203258 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone
ACCESSION AI008807
VERSION REWED56 3' end, mRNA sequence.
KEYWORDS Rattus sp.
SOURCE Rattus sp.
ORGANISM Rattus sp.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
TITLE Rattus.
JOURNAL 1 (bases 1 to 500)
COMMENT Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.

FEATURES
source
1..500
/organism="Rattus sp."
/mol_type="mRNA"
/db_xref="ATCC (inhost):2016992"
/db_xref="taxon:10118"
/clone="REMBED56"
/dev_stage="embryo 8, 12, 18 dpc"
/clone_lib="Normalized rat embryo, Bento Soares"
/note="Vector: pT7T3pac; Site_1: EcoRI; Site_2: NotI"

ORIGIN
Query Match 28.4%; Score 282.6; DB 9; Length 500;
Best Local Similarity 95.4%; Pred. No. 1e-40;
Matches 291; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 439 CTTTGTGTTCCAGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGA 498
Db 305 CCTAGTTCAGAGGACACAATGTAGGAAGCCTTTTCCACATGGCAGATGATTTGGGCAGA 246

QY 499 GCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAGGAGCAGATAAATGTTTAC 558
Db 245 GCGATGGAGTCCTTAGTATCAGTCATGATGATGATGATGATGATGATGATGATGATGAT 186

QY 559 AACTCCTGATCCCGCATGGTTTTATATATTCATATACAAAGAGGATTAGACAGTAA 618
Db 185 AACTCCTGATCCCGCATGGTTTTATATATTCATATACAAAGAGGATTAGACAGTAA 126

QY 619 GAGTTTACAGAAATAAATCTATTTTGTGAAGGTAGTGTGTTTACTGTAGATT 678
Db 125 GCGTTTACAGAAATAAATCTATTTTGTGAAGGTAGTGTGTTTACTGTAGATT 66

QY 679 CAGTAGTTCTTAACTGCTGTTTGTGTTTAACTAATGGCAGGTTTACACGCTCTATGCA 738

```

```

Db 65 CAGTAGTTCTTAACTGCTGTTTGTGTTTAACTAATGGCAGGTTTACACGCTCTATGCA 6
QY 739 ATTGT 743
Db 5 ATTGT 1

RESULT 14
LOCUS BG795230/c
DEFINITION BG795230 576 bp mRNA linear EST 16-MAY-2001
cDNA_clone UTSW_SM30E11, mRNA sequence.
ACCESSION BG795230
VERSION BG795230.1 GI:14130800
KEYWORDS Mus musculus (house mouse)
SOURCE EST.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 576)
AUTHORS Gallardo, T.D., Schageman, J.J., Pertsemilidis, A., Garner, H.R.,
UT Southwestern Medical Center, Adult Mouse Skeletal Muscle cDNA
Library
JOURNAL Unpublished (2001)
COMMENT Contact: Schageman JJ
Shohet/Garner Labs
University of Texas Southwestern Medical Center
6000 Harry Hines Blvd., NA2.226, Dallas, TX 75390, USA
Tel: 214 648 1674
Email: Jeff.Schageman@UTSouthwestern.edu
cDNA library constructed by UTSW as a component of the Program for
Genomic Applications (PGA) and the Reynolds Heart Disease
Prevention grants for use in cDNA microarray experiments. Sequence
Quality: Sequence ends were trimmed based on percentage of ambigu-
us base calls or N's in windowed segments. Sequencing: First-pass
sequencing; ABI Prism 377 sequencer and analysis software.
Seq primer: M13/pUC Reverse.

FEATURES
source
1..576
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="UTSW_SM30E11"
/sex="pooled"
/tissue_type="Diaphragm/Hind limb muscles"
/cell_type="Skeletal muscle"
/dev_stage="2 months"
/lab_host="DH5a"
/clone_lib="UTSW Adult Mouse Skeletal Muscle Library"
/note="Vector: pAMP10 (Gibco); Cloned unidirectionally.
Primer: Oligo dt. RNA isolation: cytoplasmic RNA preps
(Manniat); Cloning technique: CUA Cloning (CloneAmp,
Life Technologies); Average insert size: 1.8 Kb;
Insertion site: TAGCTCACTGAATCTGAGTG---.
Other
information regarding entire library may be found at
http://pga.swmed.edu/data/libraries/microarray_cdna_librar-
ies.htm."

ORIGIN
Query Match 28.2%; Score 281.2; DB 12; Length 576;
Best Local Similarity 95.8%; Pred. No. 1.7e-40;
Matches 297; Conservative 0; Mismatches 12; Indels 1; Gaps 1;

QY 450 AGGACACAAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGCAGACGATGGATC 509
Db 347 AGGNCNCAATGTAGGAAGCCTTTTCCACATGGCANATGATTTGGCAGACGATGGATC 288

QY 510 CTTAGTATCATGATGACATGAAGAGGAGCAGATAAATGTTTACAACTCCTGATT 569
Db 287 CTTAGTATCATGATGACATGAAGAGGAGCAGATAAATGTTTACAACTCCTGATT 238

```

```

Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
Ono, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
Shiraki, T., Tgami, M., Waki, K., Wataniki, A., Muramatsu, M. and
Hayashizaki, Y. Direct Submission
Computational Analysis of Full-length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Rnuses were provided by Michela Fagiolini and Takao K. Hensch (
Laboratory for Neuronal Circuit Development Brain Science Institute
RIKEN 2-1 Hiroawa, Wako-shi, Saitama 351-0198 Japan ) whose
assistance we gratefully acknowledge.
Please visit our web site (http://genome.sgc.riken.go.jp) for
further details.
FEATURES             Location/Qualifiers
     source            1..421
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="K430066118"
                     /tissue_type="visual cortex"
                     /clone_lib="RIKEN full-length enriched, visual cortex"
ORIGIN
Query Match      28.0%; Score 278.6; DB 13; Length 421;
Best Local Similarity 98.3%; Pred. No. 5.6e-40;
Matches 292; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
QY 450 AGGCACAATGTAGGAAGTCCTTTCCACATGGCAGATGTTGGCAGAGCGATGGAGTC 509
Db 125 AGGCACAATGTAGGAAGCCTTTCCACATGGCAGATGTTGGCAGAGCGATGGAGTC 184
QY 510 CTTAGTATCATGTCAGATGAAGAAGGAGCAGATAAATGTTTTCACTCCTGATT 569
Db 185 CTTAGTTTCAGTCATGACAGATGAAGAAGGAGCAGATAAATGTTTTCACTCCTGATT 244
QY 570 CCGCATGGTTTTATTAATTCATACACAAGAGGATTACACAGTAAGAGTTTACAAAG 629
Db 245 CCGCATGGTTTTATTAATTCATACACAAGAGGATTACACAGTAAGAGTTTACAAAG 304
QY 630 AAATCAATCATATTTTTCTGAAGGGTAGTGGTATTATCTAGTAGTTTCAGTAGTTTC 688
Db 305 AAATAAATCTATATTTTGTGAAGGGTAGTGGTACTACTAGTAGTTTCAGTAGTTTC 364
QY 689 TAAGTCCTGTTATGTTTTGTTTAAACATGGCAGGTTTTACACGCTCATGCAATTGTAC 745
Db 365 TAAGTCCTGTTATGTTTTGTTTAAACATGGCAGGTTTTACACGCTCATGCAATTGTAC 421
Search completed: April 6, 2004, 14:12:07
Job time : 3679.56 secs

```

QY	689	TAAGTCGTATTATGTTTGGTTTAAACAATGGCAGGTTTTTACACGTCCTATGCAATTGTAC	745
DB	365	TAACTCATGTTATTGTTGTTGTTTAAACAATGGCAGGTTTTTACACGTCCTATGCAATTGTAC	421

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 11:33:41 ; Search time 2486.92 Seconds  
(without alignments)  
1500.955 Million cell updates/sec

Title: US-09-966-264D-2

Perfect score: 996

Sequence: 1 GTGGTTGATTGATAGTAA.....GTGTTGATGTTAATTAATT 996

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2470632 seqs, 1873875610 residues

Total number of hits satisfying chosen parameters: 4941264

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications NA:

- 1: /cgn2\_6/ptodata/1/pubna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/1/pubna/US10C\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/1/pubna/US10\_NEW\_PUB.seq.\*
- 17: /cgn2\_6/ptodata/1/pubna/US60\_NEW\_PUB.seq.\*
- 18: /cgn2\_6/ptodata/1/pubna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	996	100.0	996	US-09-966-264-2	Sequence 2, Appli
2	410.8	41.2	5339	US-10-149-736-40	Sequence 40, Appl
3	410.8	41.2	5417	US-10-149-736-39	Sequence 39, Appl
4	410.8	41.2	5462	US-10-149-736-41	Sequence 41, Appl
5	410.8	41.2	8689	US-10-149-736-42	Sequence 42, Appl
6	410.8	41.2	11443	US-10-149-736-44	Sequence 44, Appl
7	410.8	41.2	12057	US-10-149-736-47	Sequence 47, Appl
8	410.8	41.2	13957	US-09-782-378A-22	Sequence 22, Appl
9	410.8	41.2	13957	US-09-880-107-2284	Sequence 2284, Ap
10	410.8	41.2	13957	US-10-149-736-1	Sequence 1, Appli
11	410.8	41.2	14069	US-10-342-887-434	Sequence 434, App
12	410.8	41.2	14082	US-10-342-887-981	Sequence 981, App
13	410.8	41.2	14082	US-10-341-434-108	Sequence 108, App
14	396.8	39.8	2691	US-10-149-736-38	Sequence 38, Appl
15	374.2	37.6	13815	US-10-149-736-2	Sequence 2, Appli

C 16	151.4	15.2	430	9	US-09-796-692-3505	Sequence 3505, Ap
C 17	151.4	15.2	430	14	US-10-040-862-3505	Sequence 3505, Ap
C 18	151.4	15.2	430	15	US-10-057-4758-3505	Sequence 3505, Ap
C 19	151.4	15.2	430	15	US-10-154-884B-3505	Sequence 3505, Ap
C 20	151.4	15.2	2691	15	US-10-149-736-38	Sequence 38, Appl
C 21	151.4	15.2	8689	15	US-10-149-736-42	Sequence 42, Appl
C 22	151.4	15.2	11443	15	US-10-149-736-44	Sequence 44, Appl
C 23	151.4	15.2	13957	9	US-09-782-378A-22	Sequence 22, Appl
C 24	151.4	15.2	13957	9	US-09-880-107-2284	Sequence 1, Appli
C 25	151.4	15.2	13957	15	US-10-149-736-1	Sequence 1, Appli
C 26	151.4	15.2	14069	12	US-10-342-887-434	Sequence 434, App
C 27	151.4	15.2	14082	12	US-10-342-887-981	Sequence 981, App
C 28	151.4	15.2	14082	15	US-10-341-434-108	Sequence 108, App
C 29	137	13.8	137	9	US-09-966-264-1	Sequence 1, Appli
C 30	101.4	10.2	13815	15	US-10-149-736-2	Sequence 2, Appli
C 31	51.2	5.1	3673778	14	US-10-312-841-1	Sequence 1236, Ap
C 32	50.6	5.1	12393	14	US-10-311-455-1236	Sequence 238, App
C 33	49.8	5.0	18154	14	US-10-311-455-228	Sequence 3380, Ap
C 34	49.4	5.0	622	12	US-10-424-599-3380	Sequence 176, App
C 35	48.4	4.9	23683	14	US-10-240-485-176	Sequence 364, App
C 36	48.2	4.8	5546	14	US-10-311-455-364	Sequence 36, Appl
C 37	48	4.8	8076	16	US-10-257-166-36	Sequence 418, App
C 38	47.4	4.8	5413	12	US-10-221-714A-418	Sequence 1, Appli
C 39	47.4	4.8	3673778	14	US-10-312-841-1	Sequence 2, Appli
C 40	47.4	4.8	3673778	14	US-10-312-841-2	Sequence 1455, Ap
C 41	47.2	4.7	17213	14	US-10-311-455-1455	Sequence 2075, Ap
C 42	47	4.7	4993	14	US-10-311-455-2075	Sequence 57, Appl
C 43	47	4.7	7348	12	US-10-221-714A-57	Sequence 1797, Ap
C 44	47	4.7	8771	14	US-10-311-455-1797	Sequence 1780, Ap
C 45	46.8	4.7	5683	14	US-10-311-455-1780	

ALIGNMENTS

RESULT 1  
US-09-966-264-2  
; Sequence 2, Application US/09966264  
; Patent No. US20020099015A1  
; GENERAL INFORMATION:  
; APPLICANT: Barber, Elizabeth K  
; TITLE OF INVENTION: Gene Expression Control Element DNA  
; FILE REFERENCE: 896034605001  
; CURRENT APPLICATION NUMBER: US/09/966,264  
; CURRENT FILING DATE: 2001-09-28  
; PRIOR APPLICATION NUMBER: US 60/237,079  
; PRIOR FILING DATE: 2000-09-30  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 2  
; LENGTH: 996  
; TYPE: DNA  
; ORGANISM: human  
; FEATURE:  
; NAME/KEY: exon  
; LOCATION: (1)..(996)  
; OTHER INFORMATION:  
; NAME/KEY: misc feature  
; LOCATION: (710)..(996)  
; OTHER INFORMATION: Nucleotides 710-996 are homologous to a portion of human dystroph  
; OTHER INFORMATION: in DNA in the region of exon 79 except that nucleotides 860-996 a  
; OTHER INFORMATION: re inverted in comparison to the orientation of the same sequence  
; OTHER INFORMATION: in the dystrophin DNA  
US-09-966-264-2

Query Match 100.0%; Score 996; DB 9; Length 996;

Best Local Similarity 100.0%; Pred. No. 6.9e-224; Mismatches 0; Indels 0; Gaps 0;

Matches 996; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGGTTGATTGATAGTAAATAATGTCGTTAATACAGTAGAGAGAGTAATCAAT 60

Db 1 GTGGTTGATTGATAGTAAATAATGTCGTTAATACAGTAGAGAGAGTAATCAAT 60

61 CAATCACTCATAGCAAGTGGAAAGATGATCCCATCATGGAATATCTCTGTCTGAT 120  
61 CAATCACTCATAGCAAGTGGAAAGATGATCCCATCATGGAATATCTCTGTCTGAT 120  
121 AGAATCTGTGCTATCTATGGAATCTTTTGTATATATATTTACATTTGGAACTCAAT 180  
121 AGAATCTGTGCTATCTATGGAATCTTTTGTATATATATTTACATTTGGAACTCAAT 180  
181 GTAGCTTGACATTTTCCATGTAAACACACAGTATGCTATCTATCTGACCTTTTGT 240  
181 GTAGCTTGACATTTTCCATGTAAACACACAGTATGCTATCTATCTGACCTTTTGT 240  
241 ACAACACAGTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCT 300  
241 ACAACACAGTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCT 300  
301 AATCAAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360  
301 AATCAAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360  
361 GGGATGATTTCCCAATGGCAAGAAAGACAGATGCTATCTATCTGACCTTTTGT 420  
361 GGGATGATTTCCCAATGGCAAGAAAGACAGATGCTATCTATCTGACCTTTTGT 420  
421 AAGTCTGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 480  
421 AAGTCTGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 480  
481 GCAGATGATTTGGGAGAGCGATGAGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 540  
481 GCAGATGATTTGGGAGAGCGATGAGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 540  
541 GCAGATGATTTGGGAGAGCGATGAGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 600  
541 GCAGATGATTTGGGAGAGCGATGAGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 600  
601 AAGAGGATGACAGTATGAGTATCAAGAAATATCTATTTTGTGAGGCTAGTG 660  
601 AAGAGGATGACAGTATGAGTATCAAGAAATATCTATTTTGTGAGGCTAGTG 660  
661 GTATTATCTGTAGATTTTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720  
661 GTATTATCTGTAGATTTTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720  
721 GTTTTACAGTCTATGCAATTTGACAAAAGTTTATAGAAAATCTATGTAATCTTG 780  
721 GTTTTACAGTCTATGCAATTTGACAAAAGTTTATAGAAAATCTATGTAATCTTG 780  
781 ATAGCTTAATTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840  
781 ATAGCTTAATTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840  
841 TAACAGTATTAAGAAAGATTTATAGAAAAGTTTATAGAAAATCTATGTAATCTTG 900  
841 TAACAGTATTAAGAAAGATTTATAGAAAAGTTTATAGAAAATCTATGTAATCTTG 900  
901 AGCTGTGAATCTGAGTGTGCAATTTATAGAAAAGTTTATAGAAAATCTATGTAATCTTG 960  
901 AGCTGTGAATCTGAGTGTGCAATTTATAGAAAAGTTTATAGAAAATCTATGTAATCTTG 960  
961 AATAGCATGAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 996  
961 AATAGCATGAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 996

RESULT 2  
US-10-149-736-40  
; Sequence 40, Application US/10149736  
; Publication No. US20030216332A1  
; GENERAL INFORMATION:  
; APPLICANT: Chamberlain, Jeffrey S.  
; TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences

FILE REFERENCE: UM-06968  
; CURRENT APPLICATION NUMBER: US/10/149,736  
; PRIOR FILING DATE: 2002-06-17  
; PRIOR APPLICATION NUMBER: PCT/US01/31126  
; PRIOR FILING DATE: 2001-10-04  
; PRIOR APPLICATION NUMBER: 60/238,848  
; PRIOR FILING DATE: 2000-10-06  
; NUMBER OF SEQ ID NOS: 96  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 40  
; LENGTH: 5339  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
US-10-149-736-40  
  
Query Match 41.2%; Score 410.8; DB 15; Length 5339;  
Best Local Similarity 98.3%; Pred. No. 5.7e-86;  
Matches 415; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
  
QY 450 AGGACACAATGTAGGAAGTCTTTCCACATGGCAGATGATTTGGGAGAGCGATGAGTGC 509  
DB 4491 AGGACACAATGTAGGAAGTCTTTCCACATGGCAGATGATTTGGGAGAGCGATGAGTGC 4550  
QY 510 CTTAGTATCAGTCATGACAGATGAAAGAGGAGCAGATAAATGTTTTACAATCTCTGATT 569  
DB 4551 CTTAGTATCAGTCATGACAGATGAAAGAGGAGCAGATAAATGTTTTACAATCTCTGATT 4610  
QY 570 CCGCATGCTTTTATAATATTCATCAACAAGAGGATTAGACAGTAAAGTTTACAAG 629  
DB 4611 CCGCATGCTTTTATAATATTCATCAACAAGAGGATTAGACAGTAAAGTTTACAAG 4670  
QY 630 AATAAATCTATATTTTGTGAAGGCTAGTGTATTATATCTAGATTTTCACTAGTTTCT 689  
DB 4671 AATAAATCTATATTTTGTGAAGGCTAGTGTATTATATCTAGATTTTCACTAGTTTCT 4730  
QY 690 AAGTCTGTTATGTTTGTAAACAATGGCAGGTTTACACGCTCTATGCAATTTGTCACAAA 749  
DB 4731 AAGTCTGTTATGTTTGTAAACAATGGCAGGTTTACACGCTCTATGCAATTTGTCACAAA 4790  
QY 750 AAGTTATAAAGAAATACATGTAATAATCTTTAGTAGCTAAATAACTTTGCCATTTCTTTATA 809  
DB 4791 AAGTTATAAAGAAATACATGTAATAATCTTTAGTAGCTAAATAACTTTGCCATTTCTTTATA 4850  
QY 810 TGAACGCAATTTGGTGTGTTTAAATAATTTATACAGTTTATAGAAAGAAATTTATAAGG 869  
DB 4851 TGAACGCAATTTGGTGTGTTTAAATAATTTATACAGTTTATAGAAAGAAATTTATAAGG 4910  
QY 870 AA 871  
DB 4911 AA 4912

RESULT 3  
US-10-149-736-39  
; Sequence 39, Application US/10149736  
; Publication No. US20030216332A1  
; GENERAL INFORMATION:  
; APPLICANT: Chamberlain, Jeffrey S.  
; APPLICANT: Harper, Scott Q.  
; TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences  
; FILE REFERENCE: UM-06968  
; CURRENT APPLICATION NUMBER: US/10/149,736  
; CURRENT FILING DATE: 2002-06-17  
; PRIOR APPLICATION NUMBER: PCT/US01/31126  
; PRIOR FILING DATE: 2001-10-04  
; PRIOR APPLICATION NUMBER: 60/238,848  
; PRIOR FILING DATE: 2000-10-06  
; NUMBER OF SEQ ID NOS: 96  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 39  
; LENGTH: 5417



QY 630 AATAAATCTATATTTTGTGAAGGTAGTGTATATATCTAGTATTCAGTAGTTCT 689  
 Db 6165 AATAAATCTATATTTTGTGAAGGTAGTGTATATATCTAGTATTCAGTAGTTCT 6224  
 QY 690 AAGTCGTATTGTTTGTACAAATGGCAGGTTTACACGCTATGCAATGTACAAA 749  
 Db 6225 AAGTCGTATTGTTTGTACAAATGGCAGGTTTACACGCTATGCAATGTACAAA 6284  
 QY 750 AAGTTATAAGAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCATTTCTTATA 809  
 Db 6285 AAGTTATAAGAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCATTTCTTATA 6344  
 QY 810 TGGACGCATTTGGGTTCTTTTAAATTTATAACAGTTTATAAGAAAGATTAAAGG 869  
 Db 6345 TGGACGCATTTGGGTTCTTTTAAATTTATAACAGTTTATAAGAAAGATTAAAGG 6404  
 QY 870 AA 871  
 Db 6405 AA 6406

RESULT 6  
 US-10-149-736-44  
 ; Sequence 44, Application US/10149736  
 ; Publication No. US20030216332A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Chamberlain, Jeffrey S.  
 ; APPLICANT: Harper, Scott Q.  
 ; TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences  
 ; FILE REFERENCE: UM-06968  
 ; CURRENT APPLICATION NUMBER: US/10/149,736  
 ; PRIOR FILING DATE: 2002-06-17  
 ; CURRENT APPLICATION NUMBER: PCT/US01/31126  
 ; PRIOR FILING DATE: 2000-10-04  
 ; PRIOR APPLICATION NUMBER: 60/238,848  
 ; PRIOR FILING DATE: 2000-10-06  
 ; NUMBER OF SEQ ID NOS: 96  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 44  
 ; LENGTH: 11443  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Synthetic  
 US-10-149-736-44

Query Match 41.2%; Score 410.8; DB 15; Length 11443;  
 Best Local Similarity 98.3%; Pred. No. 8.4e-86;  
 Matches 415; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 QY 450 AGGACAAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCGAGCGATGGAGTC 509  
 Db 8739 AGGACAAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCGAGCGATGGAGTC 8798  
 QY 510 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGAAATTAATGTTTACAACTCCTGATT 569  
 Db 8799 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGAAATTAATGTTTACAACTCCTGATT 8858  
 QY 570 CCGCATGTTTATATATATTCATACAAAGAGGATTAGACAGTAAGAGTTTACAAAG 629  
 Db 8859 CCGCATGTTTATATATATTCATACAAAGAGGATTAGACAGTAAGAGTTTACAAAG 8918  
 QY 630 AATAAATCTATATTTTGTGAAGGTAGTGTATATATCTAGTATTCAGTAGTTCT 689  
 Db 8919 AATAAATCTATATTTTGTGAAGGTAGTGTATATATCTAGTATTCAGTAGTTCT 8978  
 QY 690 AAGTCGTATTGTTTGTAACTAAGGCGAGTTTACACGCTATGCAATGTACAAA 749  
 Db 8979 AAGTCGTATTGTTTGTAACTAAGGCGAGTTTACACGCTATGCAATGTACAAA 9038  
 QY 750 AAGTTATAAGAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCATTTCTTATA 809  
 Db 9039 AAGTTATAAGAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCATTTCTTATA 9098

QY 810 TGGACGCATTTGGGTTCTTTTAAATTTATAACAGTTTATAAGAAAGATTAAAGG 869  
 Db 9099 TGGACGCATTTGGGTTCTTTTAAATTTATAACAGTTTATAAGAAAGATTAAAGG 9158  
 QY 870 AA 871  
 Db 9159 AA 9160  
 RESULT 7  
 US-10-149-736-47  
 ; Sequence 47, Application US/10149736  
 ; Publication No. US20030216332A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Chamberlain, Jeffrey S.  
 ; APPLICANT: Harper, Scott Q.  
 ; TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences  
 ; FILE REFERENCE: UM-06968  
 ; CURRENT APPLICATION NUMBER: US/10/149,736  
 ; CURRENT FILING DATE: 2002-06-17  
 ; PRIOR APPLICATION NUMBER: PCT/US01/31126  
 ; PRIOR FILING DATE: 2001-10-04  
 ; PRIOR APPLICATION NUMBER: 60/238,848  
 ; PRIOR FILING DATE: 2000-10-06  
 ; NUMBER OF SEQ ID NOS: 96  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 47  
 ; LENGTH: 12057  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Synthetic  
 US-10-149-736-47

Query Match 41.2%; Score 410.8; DB 15; Length 12057;  
 Best Local Similarity 98.3%; Pred. No. 8.6e-86;  
 Matches 415; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 QY 450 AGGACAAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCGAGCGATGGAGTC 509  
 Db 11214 AGGACAAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCGAGCGATGGAGTC 11273  
 QY 510 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGAAATTAATGTTTACAACTCCTGATT 569  
 Db 11274 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGAAATTAATGTTTACAACTCCTGATT 11333  
 QY 570 CCGCATGTTTATATATATTCATACAAAGAGGATTAGACAGTAAGAGTTTACAAAG 629  
 Db 11334 CCGCATGTTTATATATATTCATACAAAGAGGATTAGACAGTAAGAGTTTACAAAG 11393  
 QY 630 AATAAATCTATATTTTGTGAAGGTAGTGTATATATCTAGTATTCAGTAGTTCT 689  
 Db 11394 AATAAATCTATATTTTGTGAAGGTAGTGTATATATCTAGTATTCAGTAGTTCT 11453  
 QY 690 AAGTCGTATTGTTTGTAACTAAGGCGAGTTTACACGCTATGCAATGTACAAA 749  
 Db 11454 AAGTCGTATTGTTTGTAACTAAGGCGAGTTTACACGCTATGCAATGTACAAA 11513  
 QY 750 AAGTTATAAGAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCATTTCTTATA 809  
 Db 11514 AAGTTATAAGAAACTACATGTAAAATCTTGATAGCTAAATAACTTGCATTTCTTATA 11573  
 QY 810 TGGACGCATTTGGGTTCTTTTAAATTTATAACAGTTTATAAGAAAGATTAAAGG 869  
 Db 11574 TGGACGCATTTGGGTTCTTTTAAATTTATAACAGTTTATAAGAAAGATTAAAGG 11633  
 QY 870 AA 871  
 Db 11634 AA 11635

RESULT 8



Query Match 41.2%; Score 410.8; DB 15; Length 13957;  
 Best Local Similarity 98.3%; Pred. No. 9.3e-86;  
 Matches 415; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 450 AGGACACATGTAGGAAGTCTTTCCACATGCGCAGATGATTTGGCAGAGCGATGGAGTC 509  
 DB 11253 AGGACACATGTAGGAAGTCTTTCCACATGCGCAGATGATTTGGCAGAGCGATGGAGTC 11312  
 QY 510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGGATTAACAGTAAATGTTTACAACTCCTGATT 569  
 DB 11313 CTTAGTATCAGTCATGACAGATGAAGAAGGAGGATTAACAGTAAATGTTTACAACTCCTGATT 11372  
 QY 570 CCGCATGGTTTTTATATATTCATACACAAAGAGGATTAACAGTAAATGTTTACAACTCCTGATT 629  
 DB 11373 CCGCATGGTTTTTATATATTCATACACAAAGAGGATTAACAGTAAATGTTTACAACTCCTGATT 11432  
 QY 630 AAATAAATCTATATTTTGTGAAGGAGTGTGTTTATATCTAGTATGATTTTCAAGTATTTCT 689  
 DB 11433 AAATAAATCTATATTTTGTGAAGGAGTGTGTTTATATCTAGTATGATTTTCAAGTATTTCT 11492  
 QY 690 AAGTCTGTTATTTGTTTAAACAATGGCAGGTTTTTACAGTCTATGCAATGTACAAAA 749  
 DB 11493 AAGTCTGTTATTTGTTTAAACAATGGCAGGTTTTTACAGTCTATGCAATGTACAAAA 11552  
 QY 750 AAGTTATAAGAAAACCTACATGTAAATCTTGTAGCTAAATTAACCTTGCATTTCTTTATA 809  
 DB 11553 AAGTTATAAGAAAACCTACATGTAAATCTTGTAGCTAAATTAACCTTGCATTTCTTTATA 11612  
 QY 810 TGAACGCGATTTGGTGTGTTTAAAAATTTATAACAGTTATTAAGAAAAGATTGTAACCTA 869  
 DB 11613 TGAACGCGATTTGGGTTGTTTAAAAATTTATAACAGTTATTAAGAAAAGATTGTAACCTA 11672  
 QY 870 AA 871  
 DB 11673 AA 11674

RESULT 11  
 US-10-342-887-434  
 ; Sequence 434, Application US/10342887  
 ; Publication No. US20040058340A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dai, Hongyue  
 ; APPLICANT: He, Yudong  
 ; APPLICANT: Linsley, Peter S.  
 ; APPLICANT: Mac, Mao  
 ; APPLICANT: Roberts, Christopher J.  
 ; APPLICANT: Van 't Veer, Laura Johanna  
 ; APPLICANT: Van de Vijver, Marc J.  
 ; APPLICANT: Bernards, Rene  
 ; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients  
 ; FILE REFERENCE: 9301-188-999  
 ; CURRENT APPLICATION NUMBER: US/10/342,887  
 ; CURRENT FILING DATE: 2003-01-15  
 ; PRIOR APPLICATION NUMBER: 60/298,918  
 ; PRIOR FILING DATE: 2001-06-18  
 ; PRIOR APPLICATION NUMBER: 60/380,710  
 ; PRIOR FILING DATE: 2002-05-14  
 ; PRIOR APPLICATION NUMBER: 10/172,118  
 ; PRIOR FILING DATE: 2002-06-14  
 ; NUMBER OF SEQ ID NOS: 2699  
 ; SEQ ID NO 434  
 ; LENGTH: 14069  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-10-342-887-434

Query Match 41.2%; Score 410.8; DB 12; Length 14069;  
 Best Local Similarity 98.3%; Pred. No. 9.3e-86;  
 Matches 415; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 450 AGGACACATGTAGGAAGTCTTTCCACATGCGCAGATGATTTGGCAGAGCGATGGAGTC 509

DB 11365 AGGACACATGTAGGAAGTCTTTCCACATGCGCAGATGATTTGGCAGAGCGATGGAGTC 11424  
 QY 510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGGATTAACAGTAAATGTTTACAACTCCTGATT 569  
 DB 11425 CTTAGTATCAGTCATGACAGATGAAGAAGGAGGATTAACAGTAAATGTTTACAACTCCTGATT 11484  
 QY 570 CCGCATGGTTTTTATATATTCATACACAAAGAGGATTAACAGTAAATGTTTACAACTCCTGATT 629  
 DB 11485 CCGCATGGTTTTTATATATTCATACACAAAGAGGATTAACAGTAAATGTTTACAACTCCTGATT 11544  
 QY 630 AAATAAATCTATATTTTGTGAAGGAGTGTGTTTATATCTAGTATGATTTTCAAGTATTTCT 689  
 DB 11545 AAATAAATCTATATTTTGTGAAGGAGTGTGTTTATATCTAGTATGATTTTCAAGTATTTCT 11604  
 QY 690 AAGTCTGTTATTTGTTTAAACAATGGCAGGTTTTTACAGTCTATGCAATGTACAAAA 749  
 DB 11605 AAGTCTGTTATTTGTTTAAACAATGGCAGGTTTTTACAGTCTATGCAATGTACAAAA 11664  
 QY 750 AAGTTATAAGAAAACCTACATGTAAATCTTGTAGCTAAATTAACCTTGCATTTCTTTATA 809  
 DB 11665 AAGTTATAAGAAAACCTACATGTAAATCTTGTAGCTAAATTAACCTTGCATTTCTTTATA 11724  
 QY 810 TGAACGCGATTTGGTGTGTTTAAAAATTTATAACAGTTATTAAGAAAAGATTGTAACCTA 869  
 DB 11725 TGAACGCGATTTGGGTTGTTTAAAAATTTATAACAGTTATTAAGAAAAGATTGTAACCTA 11784  
 QY 870 AA 871  
 DB 11785 AA 11786

RESULT 12  
 US-10-342-887-981  
 ; Sequence 981, Application US/10342887  
 ; Publication No. US20040058340A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dai, Hongyue  
 ; APPLICANT: He, Yudong  
 ; APPLICANT: Linsley, Peter S.  
 ; APPLICANT: Mac, Mao  
 ; APPLICANT: Roberts, Christopher J.  
 ; APPLICANT: Van 't Veer, Laura Johanna  
 ; APPLICANT: Van de Vijver, Marc J.  
 ; APPLICANT: Bernards, Rene  
 ; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients  
 ; FILE REFERENCE: 9301-188-999  
 ; CURRENT APPLICATION NUMBER: US/10/342,887  
 ; CURRENT FILING DATE: 2003-01-15  
 ; PRIOR APPLICATION NUMBER: 60/298,918  
 ; PRIOR FILING DATE: 2001-06-18  
 ; PRIOR APPLICATION NUMBER: 60/380,710  
 ; PRIOR FILING DATE: 2002-05-14  
 ; PRIOR APPLICATION NUMBER: 10/172,118  
 ; PRIOR FILING DATE: 2002-06-14  
 ; NUMBER OF SEQ ID NOS: 2699  
 ; SEQ ID NO 981  
 ; LENGTH: 14082  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-10-342-887-981

Query Match 41.2%; Score 410.8; DB 12; Length 14082;  
 Best Local Similarity 98.3%; Pred. No. 9.4e-86;  
 Matches 415; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 450 AGGACACATGTAGGAAGTCTTTCCACATGCGCAGATGATTTGGCAGAGCGATGGAGTC 509  
 DB 11378 AGGACACATGTAGGAAGTCTTTCCACATGCGCAGATGATTTGGCAGAGCGATGGAGTC 11437  
 QY 510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGGATTAACAGTAAATGTTTACAACTCCTGATT 569  
 DB 11438 CTTAGTATCAGTCATGACAGATGAAGAAGGAGGATTAACAGTAAATGTTTACAACTCCTGATT 11497





```

; Sequence 2, Application US/10149736
; Publication No. US20030216332A1
; GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences
; FILE REFERENCE: UM-06968
; CURRENT APPLICATION NUMBER: US/10/149,736
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: PCT/US01/31126
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,848
; PRIOR FILING DATE: 2000-10-06
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 13815
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-149-736-2

```

```

Query Match      37.6%; Score 374.2; DB 15; Length 13815;
Best Local Similarity 96.0%; Pred. No. 3.7e-77;
Matches 406; Conservative 0; Mismatches 13; Indels 4; Gaps 2;

QY 450 AGGACACAATGTAGGAAGTCTTTCCACATGCGCAGATGTTGGCAGAGCGATGGAGTC 509
DB 11244 AGGACACAATGTAGGAAGCCTTTCCACATGCGCAGATGTTGGCAGAGCGATGGAGTC 11303

QY 510 CTTAGTATCAGTCATGACAGATGAGAGGAGGACAGATAAATGTTTACAACTCCTGATT 569
DB 11304 CTTAGTTCAGTCATGACAGATGAGAGGAGGACAGATAAATGTTTACAACTCCTGATT 11363

QY 570 CCCGCATGCTTTTATATATTTATATATTTATATATTTATATATTTATATATTTATATATTT 629
DB 11364 CCCGCATGCTTTTATATATTTATATATTTATATATTTATATATTTATATATTTATATATTT 11423

QY 630 AAAT-AAATCTATATTTTGTGAAGGAGTAGTGTATATATATCTAGTATGATTTTCAGTAGTTTC 688
DB 11424 AAATAAATCTATATTTTGTGAAGGAGTAGTGTATATATCTAGTATGATTTTCAGTAGTTTC 11483

QY 689 TAAGTCTGTTATTGTTTGTAACTGCGAGGTTTACAGTCTATGCAATTTGTACAAA 748
DB 11484 TAAGTCTGTTATTGTTTGTAACTGCGAGGTTTACAGTCTATGCAATTTGTACAAA 11543

QY 749 AAAGTTATAAGAAACTACATGTAATAATCTTGATAGCTAAATAACTTCCCATTTCTTTAT 808
DB 11544 AAAGTTAAAGAAA---ACATGTAATAATCTTGATAGCTAAATAACTTCCCATTTCTTTAT 11600

QY 809 ATGGAACGCATTTGGGTTGTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 868
DB 11601 ATGGAACGCATTTGGGTTGTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAACT 11660

QY 869 GAA 871
DB 11661 AAA 11663

```

Search completed: April 6, 2004, 15:01:46  
Job time : 2493.92 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 14:12:12 ; Search time 3672.8 Seconds

(without alignments)  
8098.097 Million cell updates/sec

Title: US-09-966-264D-2

Perfect score: 996

Sequence: 1 ggggttgatgatgataaa.....gtgttgatgtaatt 996

Scoring table: OLIGO.NUC

Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 0

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

1: em\_estba.\*

2: em\_esthum.\*

3: em\_estin.\*

4: em\_estmu.\*

5: em\_estov.\*

6: em\_estpl.\*

7: em\_estro.\*

8: em\_htc.\*

9: gb\_est1.\*

10: gb\_est2.\*

11: gb\_hcc.\*

12: gb\_est3.\*

13: gb\_est4.\*

14: gb\_est5.\*

15: em\_estfun.\*

16: em\_estom.\*

17: em\_gss\_hum.\*

18: em\_gss\_inv.\*

19: em\_gss\_pln.\*

20: em\_gss\_vrt.\*

21: em\_gss\_fun.\*

22: em\_gss\_mam.\*

23: em\_gss\_mus.\*

24: em\_gss\_pro.\*

25: em\_gss\_rtd.\*

26: em\_gss\_phg.\*

27: em\_gss\_vrl.\*

28: gb\_gsl1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Query %	Length	ID	Description
1	410	41.2	844	12	BG706268	BG706268 602669567
2	342	34.3	629	9	AL712236	AL712236 DKFZp686J
3	342	34.3	629	13	BX509925	BX509925 DKFZp686J
4	238	23.9	999	13	BX419180	BX419180 BX419180

5	231	23.2	739	9	AUI38648	AUI38648 AUI38648
6	221	22.2	308	14	N75050	N75050 za89C07.61
7	209	21.0	275	10	AW604756	AW604756 CM0-CT034
8	209	21.0	277	10	AW861957	AW861957 CM3-CT034
9	209	21.0	284	10	AW858457	AW858457 CM3-CT034
10	208	20.9	888	14	CF552038	CF552038 AGENCOURT
11	202	20.3	525	28	AQ579243	AQ579243 HS 5417.A
12	194	19.5	271	10	AW858578	AW858578 CM3-CT034
13	183	18.4	619	10	AW771158	AW771158 hn53c06.x
14	161	16.2	309	14	H27701	H27701 Y161C12.r1
15	148	14.9	556	28	AQ504173	AQ504173 RPCI-11-2
16	145	14.7	396	14	H89576	H89576 YW28C12.r1
17	145	14.7	455	9	AA427831	AA427831 zw49412.r
18	146	14.7	482	14	CB045405	CB045405 NISC_gcl0
19	146	14.7	473	14	CA389721	CA389721 cs101f03
20	146	14.7	489	28	AQ015185	AQ015185 CIT-HSP-2
21	146	14.7	647	10	AW664684	AW664684 h184e10.x
22	146	14.7	696	10	AW950480	AW950480 EST362550
23	146	14.7	710	9	AV725574	AV725574 AV725574
24	146	14.7	727	12	BG567176	BG567176 60289640
25	146	14.7	740	14	CD356811	CD356811 AGENCOURT
26	146	14.7	745	14	CD357556	CD357556 AGENCOURT
27	146	14.7	1121	12	BM546012	BM546012 AGENCOURT
28	145	14.6	743	14	CD110642	CD110642 AGENCOURT
29	141	14.2	373	10	AW580404	AW580404 PM2-HT045
30	129	13.0	939	13	BO718541	BO718541 AGENCOURT
31	120	12.0	352	10	AW385154	AW385154 PM2-HT045
32	118	11.8	501	13	HU619605	HU619605 UI-H-FH1-
33	112	11.2	862	14	CB962272	CB962272 AGENCOURT
34	111	11.1	353	10	AW580423	AW580423 PM2-HT045
35	109	10.9	183	10	AW858537	AW858537 CM3-CT034
36	97	9.7	911	12	BI752714	BI752714 603028327
37	94	9.4	439	14	CB750007	CB750007 AMGNNUC.S
38	94	9.4	924	13	BQ327942	BQ327942 AGENCOURT
39	94	9.4	4437	11	AK036936	AK036936 Mus muscu
40	91	9.1	492	14	CB725360	CB725360 AMGNNUC.N
41	91	9.1	571	14	CA893354	CA893354 B0178E05-
42	90	9.0	375	10	AW607064	AW607064 PM2-HT045
43	88	8.8	484	10	BF655326	BF655326 279797.MA
44	87	8.7	531	12	BI281078	BI281078 UI-R-DD0-
45	87	8.7	576	12	BG795230	BG795230 UTSW_SM30

## ALIGNMENTS

RESULT 1  
BG706268  
LOCUS 602669567F1 NIH\_MGC\_96 Homo sapiens CDNA clone IMAGE:4792313 5',  
DEFINITION 844 bp mRNA linear EST 07-MAY-2001  
mRNA sequence.  
ACCESSION BG706268  
VERSION BG706268.1 GI:13981445  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 844)  
AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-r@mail.nih.gov](mailto:cgabs-r@mail.nih.gov)  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: LLAM10670 row: e column: 18

High quality sequence stop: 658.  
 Location/Qualifiers  
 1. .844  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4792313"  
 /tissue\_type="hypothalamus"  
 /lab\_host="DH10B"  
 /clone\_lib="NIH\_MGC\_96"  
 /note="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (Gtcag); Oligo-dT primed using primer 5'-TTTTTCTTTTCTTTT-3', size-selected for average insert size 2.3 kb and normalized to R0T 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH\_MGC Library."

ORIGIN  
 Query Match 41.2%; Score 410; DB 12; Length 844;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-189;  
 Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 450 AGGACACATGATAGGAGCTTTTCCACATGCGCAGATGTTGGCAGAGCGATGGAGTC 509  
 |||||  
 Db 88 AGGACACAAATGAGGAAGTCTTTCCATGTCGATGATTTGGCAGAGCGATGGAGTC 147  
 |||||

QY 510 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGATAAATGTTTCAACCTCCTGATT 569  
 |||||  
 Db 148 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGATAAATGTTTCAACCTCCTGATT 207  
 |||||

QY 570 CCGGATGTTTTTATAATTCATACAAAGAGGATTAGACAGTAGAGTTTACAAAG 629  
 |||||  
 Db 208 CCGGATGTTTTTATAATTCATACAAAGAGGATTAGACAGTAGAGTTTACAAAG 267  
 |||||

QY 630 AAATAAATCTATATTTCTGGAAGGTTAGTGTATTATCTAGATTTTCAAGTAGTTCT 689  
 |||||  
 Db 268 AAATAAATCTATATTTCTGGAAGGTTAGTGTATTATCTAGATTTTCAAGTAGTTCT 327  
 |||||

QY 690 AAGTCGTGTTATTTGTTTAAATGAGGAGGTTTACACGTCTATGCAATTTGTCACAAA 749  
 |||||  
 Db 328 AAGTCGTGTTATTTGTTTAAATGAGGAGGTTTACACGTCTATGCAATTTGTCACAAA 387  
 |||||

QY 750 AAGTTATAGAAAACCTACATGTAATCTTGATAGCTAAATTAATGCAATTTCTTTATA 809  
 |||||  
 Db 388 AAGTTATAGAAAACCTACATGTAATCTTGATAGCTAAATTAATGCAATTTCTTTATA 447  
 |||||

QY 810 TGAACGCAATTTGGGTTGTTTAAAAATTTATAACAGTTTATAAGAAAGA 859  
 |||||  
 Db 448 TGAACGCAATTTGGGTTGTTTAAAAATTTATAACAGTTTATAAGAAAGA 497  
 |||||

RESULT 2  
 AL712236 629 bp mRNA linear EST 04-SEP-2003  
 LOCUS DKFZp686J0287 r1 686 (synonym: hlcc3) Homo sapiens cDNA clone  
 DEFINITION DKFZp686J0287 5', mRNA sequence.  
 ACCESSION AL712236  
 VERSION AL712236.2 GI:32026719  
 EST.  
 KEYWORDS Homo sapiens (human)  
 SOURCE  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 629)  
 Poustka,A., Albert,R., Moosmayer,P., Schupp,I., Wellenreuther,R.,  
 Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and  
 Wiemann,S.  
 TITLE (Poustka,A., Albert,R., Moosmayer,P., Schupp,I.,  
 Wellenreuther,R., et al.)

Unpublished (2003)  
 On Mar 22, 2002 this sequence version replaced gi:19695591.  
 Contact: MIPS  
 MIPS  
 Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany  
 This is the 5' sequence of the clone insert  
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
 Research Center (DKFZ); Email s.wiemann@kfz-heidelberg.de;  
 sequenced by DKFZ (German Cancer Research Center,  
 Heidelberg/Germany) within the cDNA sequencing consortium of the  
 German Genome Project.  
 No sl sequence available  
 This clone (DKFZp686J0287) is available at the RZPD in Berlin.  
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.  
 Location/Qualifiers  
 1. .629  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="DKFZp686J0287"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /clone\_lib="686 (synonym: hlcc3)"  
 /note="Vector: pTriplex2; Site\_1: SfiIA; Site\_2: SfiIB;  
 cDNA-collection"

FEATURES  
 source

ORIGIN  
 Query Match 34.3%; Score 342; DB 9; Length 629;  
 Best Local Similarity 100.0%; Pred. No. 3e-156;  
 Matches 342; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 518 CAGTCATGACAGATGAAGAGGAGCAGATAAATGTTTACACTCCTGATCCCCCATG 577  
 |||||  
 Db 1 CAGTCATGACAGATGAAGAGGAGCAGATAAATGTTTACACTCCTGATCCCCCATG 60  
 |||||

QY 578 GTTTTTATATATTCATACAAAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAAT 637  
 |||||  
 Db 61 GTTTTTATATATTCATACAAAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAAT 120  
 |||||

QY 638 CTATATTTTGTGAAGGTTAGTGTATTATCTAGTAGATTTCAAGTCTCTTAAAGTCTGT 697  
 |||||  
 Db 121 CTATATTTTGTGAAGGTTAGTGTATTATCTAGTAGATTTCAAGTCTCTTAAAGTCTGT 180  
 |||||

QY 698 TATTGTTTCTTAAACATGCGAGGTTTACACGCTCTATGCAATTTGACAAAAAGTTATA 757  
 |||||  
 Db 181 TATTGTTTCTTAAACATGCGAGGTTTACACGCTCTATGCAATTTGACAAAAAGTTATA 240  
 |||||

QY 758 AGAAACATCATGTAAATCTTGATAGCTAAATAACTTGCATTTCTTTATATGGAACGC 817  
 |||||  
 Db 241 AGAAACATCATGTAAATCTTGATAGCTAAATAACTTGCATTTCTTTATATGGAACGC 300  
 |||||

QY 818 ATTTTGGGTTGTTTAAAAATTTATAACAGTTTATAAGAAAGA 859  
 |||||  
 Db 301 ATTTTGGGTTGTTTAAAAATTTATAACAGTTTATAAGAAAGA 342  
 |||||

RESULT 3  
 BX509925 629 bp mRNA linear EST 04-SEP-2003  
 LOCUS DKFZp686J2487 r1 686 (synonym: hlcc3) Homo sapiens cDNA clone  
 DEFINITION DKFZp686J2487 5', mRNA sequence.  
 ACCESSION BX509925  
 VERSION BX509925.1 GI:32049928  
 EST.  
 KEYWORDS Homo sapiens (human)  
 SOURCE  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 629)  
 Poustka,A., Albert,R., Moosmayer,P., Schupp,I., Wellenreuther,R.,  
 Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and  
 Wiemann,S.  
 REFERENCE  
 AUTHORS

EST (Poustka,A., Albert,R., Moosmayer,P., Schupp,I.,  
Wellenreuther,R., et al.)  
Unpublished (2003)  
Contact: MIPS

INGOLSTAEDTER Landstr.1, D-85764 Neuherberg, Germany  
This is the 5' sequence of the clone insert  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
sequenced by DKFZ (German Cancer Research Center,  
Heidelberg/Germany) within the cDNA sequencing consortium of the  
German Genome Project.  
No sl sequence available.  
This clone (DKFZ686J2487) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.  
Location/Qualifiers  
1. .629  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="DKFZ686J2487"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/clone\_lib="686" (synonym: hlcc3)"  
/note="Vector: pTriplex2, Site\_1: SfIIA; Site\_2: SfiIB;  
cDNA-collection"

ORIGIN  
Query Match 34.3%; Score 342; DB 13; Length 629;  
Best Local Similarity 100.0%; Pred. No. 3e-156;  
Matches 342; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 518 CAGTCATGACAGATGACGAGGAGCAGTAATGTTTACAACTCTGATCCCGCATG 577  
Db 1 CAGTCATGACAGATGACGAGGAGCAGTAATGTTTACAACTCTGATCCCGCATG 60

QY 578 GTTTTATAATTCATACACAAAGAGGATTAGACAGTAGAGTTTACAGAAATAAAT 637  
Db 61 GTTTTATAATTCATACACAAAGAGGATTAGACAGTAGAGTTTACAGAAATAAAT 120

QY 638 CTATATTTTGTGAGGAGGTAGTATATATCTAGATTTCAGTAGTTCTTAAGTCTGT 697  
Db 121 CTATATTTTGTGAGGAGGTAGTATATATCTAGATTTCAGTAGTTCTTAAGTCTGT 180

QY 698 TATTGTTTGTAACTAGGAGGTTTACACGCTCTATGCAATGTACAAAGAGTTATA 757  
Db 181 TATTGTTTGTAACTAGGAGGTTTACACGCTCTATGCAATGTACAAAGAGTTATA 240

QY 758 AGAAACTACATGTAAATCTGTAGCTAAATAACTGGCCATTTCTTTATATGGAACG 817  
Db 241 AGAAACTACATGTAAATCTGTAGCTAAATAACTGGCCATTTCTTTATATGGAACG 300

QY 818 ATTTGGTGTGTTTAAATTTATAACAGTGTATAAGAAAGA 859  
Db 301 ATTTGGTGTGTTTAAATTTATAACAGTGTATAAGAAAGA 342

RESULT 4  
BX419180  
LOCUS  
DEFINITION  
BX419180 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone  
CS0DF014YA17 5-PRIME, mRNA sequence.  
ACCESSION  
BX419180  
VERSION  
BX419180.1 GI:30637926  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 999)  
Li,W.B., Gruber,C., Jesse,J., and Polayes,D.  
Full-length cDNA libraries and normalization

Unpublished (2001)  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 10533.r For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0DF014AA090P1&cluster=10533.r. Contact :  
Peng Liang Email: fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0DF014AA09Q1.  
Location/Qualifiers  
1. .999  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DF014YA17"  
/tissue\_type="FETAL BRAIN"  
/dev\_stage="fetal"  
/clone\_lib="Homo sapiens FETAL BRAIN"  
/note="Origin: brain; Vector: pCMVSPORT 6; 1st strand cDNA  
was primed with a NotI-oligo(dT) primer. Five prime end  
enriched double-strand cDNA was digested with Not I and  
cloned into the Not I and EcoRV sites of the pCMVSPORT 6  
vector. Library was not normalized."

ORIGIN  
Query Match 23.9%; Score 238; DB 13; Length 999;  
Best Local Similarity 100.0%; Pred. No. 2.4e-105;  
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 622 TTTCAGAAATAAATCTATATTTTGTGCAAGGAGTAGTGATTTATATACGTGATTCAG 681  
Db 69 TTTCAGAAATAAATCTATATTTTGTGCAAGGAGTAGTGATTTATATACGTGATTCAG 128

QY 682 TAGTTTCTAAGTCTGTATTGTTTGTAAACAATGCGAGGTTTACACGTCTATGCAATT 741  
Db 129 TAGTTTCTAAGTCTGTATTGTTTGTAAACAATGCGAGGTTTACACGTCTATGCAATT 188

QY 742 GTACAAAAAGTTATAAGAAAACTACATGTAATAAATCTTGATAGCTAAATAACTTCCATT 801  
Db 189 GTACAAAAAGTTATAAGAAAACTACATGTAATAAATCTTGATAGCTAAATAACTTCCATT 248

QY 802 TCTTTATATGGAACGCAATTTGGTGTGTTTAAATAATTTATACAGTTATAAGAAAGA 859  
Db 249 TCTTTATATGGAACGCAATTTGGTGTGTTTAAATAATTTATACAGTTATAAGAAAGA 306

RESULT 5  
AU138648  
LOCUS  
DEFINITION  
AU138648 PLACE1 Homo sapiens cDNA clone PLACE1009015 5', mRNA  
sequence.  
ACCESSION  
AU138648  
VERSION  
AU138648.1 GI:11000169  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 739)  
Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,  
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and  
Isogai,T.  
TITLE  
HRI human cDNA project  
JOURNAL  
Unpublished (2000)  
CONTACT: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3975

Fax: 81-438-52-3986  
Email: genomics@ri.co.jp  
ERI human cDNA project; 5'- & 3'-end one pass sequencing: Helix  
Research Institute; cDNA library construction: Department of  
Virology, Institute of Medical Science, University of Tokyo, and  
Helix Research Institute.  
Location/Qualifiers  
1. 739

## FEATURES

source

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="PLACE1009015"  
/tissue\_type="placenta"  
/clone\_lib="PLACE1"  
/note="Vector: pME18SFL3"

## ORIGIN

Query Match 23.2%; Score 231; DB 9; Length 739;  
Best Local Similarity 100.0%; Pred. NO. 6.8e-102;  
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 629 GAAATAAATCTATATTTTGTGAAGGAGTAGTGGTATTATACGTAGATTTCAGTAGTTTC 688  
Db 38 GAAATAAATCTATATTTTGTGAAGGAGTAGTGGTATTATACGTAGATTTCAGTAGTTTC 97  
QY 689 TAAGTCGTATGTTTGTGTTTAAACATGCGAGGTTTACACGCTCTATGCAATGTCACAAA 748  
Db 98 TAAGTCGTATGTTTGTGTTTAAACATGCGAGGTTTACACGCTCTATGCAATGTCACAAA 157  
QY 749 AAGATTAAGAACTACATGTAATCTTTCATAGCTTAATAAATTCGCAATTCCTTTAT 808  
Db 158 AAGATTAAGAACTACATGTAATCTTTCATAGCTTAATAAATTCGCAATTCCTTTAT 217  
QY 809 ATGGAACGCAATTTGGTGGTTTAAATAATTTTAAACAGTTTATAAAGAAAGA 859  
Db 218 ATGGAACGCAATTTGGTGGTTTAAATAATTTTAAACAGTTTATAAAGAAAGA 268

## RESULT 6

N75050/c  
LOCUS N75050 308 bp mRNA linear EST 29-MAR-1996  
DEFINITION za89c07.s1 Soares fetal lung NBHL19W Homo sapiens cDNA clone  
IMAGE:299724 3' similar to gb:M18533 DYSTROPHIN (HUMAN);, mRNA  
sequence.  
N75050 GI:1237628  
VERSION EST.  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens

## REFERENCE

1 (bases 1 to 308)  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
Trevisan, S., Waterston, R., Williamson, A., Wohlmann, P. and  
Wilson, R.  
The WashU-Merck EST Project  
Unpublished (1995)  
CONTACT: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (infoimage.lnl.gov) for further information.  
Seq primer: ml3 -40 forward  
High quality sequence stop: 227.

## TITLE

## JOURNAL

## COMMENT

## FEATURES

source

1. 308  
/organism="Homo sapiens"  
/mol\_type="mRNA"

/db\_xref="GDB:1244648"  
/db\_xref="taxon:9606"  
/clone="IMAGE:299724"  
/dev\_stage="19 weeks"  
/lab\_hosts="DH10B (ampicillin resistant)"  
/clone\_lib="Soares fetal lung NBHL19W"  
/notes="Organ: lung; Vector: pF73D (Pharmacia) with a  
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer  
[5'-TGTTACCAATCTCAAGTGGAGCGCGCAATTTTCTTTT-3'],  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pF73 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Bento  
Soares and M.Fatima Bonaldo. This library was constructed  
from the same fetus as the fetal heart library, Soares  
fetal heart NBHL19W."

## ORIGIN

Query Match 22.2%; Score 221; DB 14; Length 308;  
Best Local Similarity 100.0%; Pred. NO. 6.2e-97;  
Matches 221; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 639 TATATTTTGTGAAGGAGTAGTGGTATTATACGTAGATTTCAGTAGTTTCAGTAGTTTC 698  
Db 230 TATATTTTGTGAAGGAGTAGTGGTATTATACGTAGATTTCAGTAGTTTCAGTAGTTTC 171  
QY 699 ATTGTTTGTTAACAATGCGAGGTTTACACGCTCTATGCAATGTCACAAAAGTTATAA 758  
Db 170 ATTGTTTGTTAACAATGCGAGGTTTACACGCTCTATGCAATGTCACAAAAGTTATAA 111  
QY 759 GAAACTACATGTAAATCTTGTAGTAACTAAATTAACCTTGCATTCTTTATATGGAACGCA 818  
Db 110 GAAACTACATGTAAATCTTGTAGTAACTAAATTAACCTTGCATTCTTTATGGAACGCA 51  
QY 819 TTTTGGTGTGTTTAAAAATTTTAAACAGTTTATAAAGAAAGA 859  
Db 50 TTTTGGTGTGTTTAAAAATTTTAAACAGTTTATAAAGAAAGA 10

## RESULT 7

N75050/c  
LOCUS N75050 275 bp mRNA linear EST 23-MAR-2000  
DEFINITION CM0-CT0341-260100-160-a10 CT0341 Homo sapiens cDNA, mRNA sequence.  
IMAGE:299724 3' similar to gb:M18533 DYSTROPHIN (HUMAN);, mRNA  
sequence.  
N75050 GI:7309497  
VERSION EST.  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens

## REFERENCE

1 (bases 1 to 275)  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
HCGP http://www.ludwig.org.br/ORESTES.  
The FAPESP/LICR Human Cancer Genome Project  
Unpublished (1999)  
CONTACT: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM0&t2=CM0-CT0341-  
260100-160-a10&t3=2000-01-36&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 10  
High quality sequence stop: 275.  
Location/Qualifiers  
1. 275

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="CT0341"
/notes="Organ: colon; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN
Query Match      21.0%; Score 209; DB 10; Length 275;
Best Local Similarity 100.0%; Pred. No. 4.8e-91;
Matches 209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 450 AGACACAAATGAGGAAGTCCTTTCCACATGGCAGATGTTGGCAGAGCGATGGAGTC 509
DB 55 AGACACAAATGAGGAAGTCCTTTCCACATGGCAGATGTTGGCAGAGCGATGGAGTC 114
QY 510 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGATAAATGTTTACAACTCCTGATT 569
DB 115 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGATAAATGTTTACAACTCCTGATT 174
QY 570 CCCGATGCTTTTATAATTTATATTCATACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
DB 175 CCCGATGCTTTTATAATTTATATTCATACAAAGAGGATTAGACAGTAAGAGTTTACAAG 234
QY 630 AAATAAATCTATATTTTGTGAAGGGTAG 658
DB 235 AAATAAATCTATATTTTGTGAAGGGTAG 263

RESULT 8
AW861957
LOCUS
DEFINITION
CM3-CT0341-190400-152-c06 CT0341 Homo sapiens cDNA, mRNA sequence.
ACCESSION
AW861957.1 GI:7957650
VERSION
EST.
KEYWORDS
Homo sapiens (human)
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 277)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jorgeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
10737800
PUBMED
JOURNAL
MEDLINE
COMMENT
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-CM3-CT0341-190
400-152-c06&t3=2000-04-19&t4=1)
400-152-c06&t3=2000-04-19&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 12
High quality sequence stop: 277.

```

```

FEATURES
source
Location/Qualifiers
1..277
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="CT0341"
/notes="Organ: colon; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN
Query Match      21.0%; Score 209; DB 10; Length 277;
Best Local Similarity 100.0%; Pred. No. 4.8e-91;
Matches 209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 450 AGACACAAATGAGGAAGTCCTTTCCACATGGCAGATGTTGGCAGAGCGATGGAGTC 509
DB 57 AGACACAAATGAGGAAGTCCTTTCCACATGGCAGATGTTGGCAGAGCGATGGAGTC 116
QY 510 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGATAAATGTTTACAACTCCTGATT 569
DB 117 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGATAAATGTTTACAACTCCTGATT 176
QY 570 CCCGATGCTTTTATAATTTATATTCATACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
DB 177 CCCGATGCTTTTATAATTTATATTCATACAAAGAGGATTAGACAGTAAGAGTTTACAAG 236
QY 630 AAATAAATCTATATTTTGTGAAGGGTAG 658
DB 237 AAATAAATCTATATTTTGTGAAGGGTAG 265

RESULT 9
AW858457
LOCUS
DEFINITION
CM3-CT0341-170200-093-b02 CT0341 Homo sapiens cDNA, mRNA sequence.
ACCESSION
AW858457
VERSION
AW858457.1 GI:7954150
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 284)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jorgeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
10737800
PUBMED
JOURNAL
MEDLINE
COMMENT
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-CM3-CT0341-170
200-093-b02&t3=2000-02-17&t4=1)
200-093-b02&t3=2000-02-17&t4=1)
Seq primer: puc 18 forward

```

High quality sequence stop: 284.  
Location/Qualifiers  
1. .284  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="CT0341"  
/note="Organ: colon; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 21.0%; Score 209; DB 10; Length 284;  
Best Local Similarity 100.0%; Pred. No. 4.7e-91;  
Matches 209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 450 AGGACACATGAGGAGGCTTTCCACATGGCAGATGTTGGCAGAGGATGGAGTC 509  
DB- 64 AGGACACATGAGGAGGCTTTCCACATGGCAGATGTTGGCAGAGGATGGAGTC 123  
QY 510 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGATAAATGTTTACAACTCCTGATT 569  
DB 124 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGATAAATGTTTACAACTCCTGATT 183  
QY 570 CCGGATGTTTTTATATATTCATACACAAAGAGGATTACAGATGAAGTTTACAA 629  
DB 184 CCGGATGTTTTTATATATTCATACACAAAGAGGATTACAGATGAAGTTTACAA 243  
QY 630 AAATAAATCTATATTTTCTGGAAGGCTAG 658  
DB 244 AAATAAATCTATATTTCTGGAAGGCTAG 272

RESULT 10  
CF552038 888 bp mRNA linear EST 22-SEP-2003  
LOCUS  
DEFINITION AGENCOURT 15595498 NIH.MGC 183 Homo sapiens cDNA clone  
IMAGE:30529933 5', mRNA sequence.  
ACCESSION CF552038.1 GI:34888872  
VERSION EST.  
KEYWORDS Homo sapiens (human)  
SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 888)  
NIH-MGC <http://mgs.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Daniela S. Gerhard, Ph.D.  
Office of Cancer Genomics  
National Cancer Institute / NIH  
Bldg. 31 Rm10A07 Bethesda, MD 20892  
Email: [cgabs@mail.nih.gov](mailto:cgabs@mail.nih.gov)  
Tissue Procurement: Dr. Michael Brownstein  
cDNA Library Preparation: Invitrogen Corp  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: NDAM618 row: e column: 14  
High quality sequence start: 17  
High quality sequence stop: 708.  
Location/Qualifiers  
1. .888  
/organism="Homo sapiens"  
/mol\_type="mRNA"

FEATURES  
source

/db\_xref="taxon:9606"  
/clone="IMAGE:30529933"  
/lab\_host="DH10B-Tona (T1 and T5 phage resistant)"  
/clone\_lib="NIH MGC 183"  
/note="Organ: Pooled muscle (cardiac and skeletal); Vector: pCMV-SPORT6.1; Site 1: EcoRV (destroyed); Site 2: NotI; Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.7. Library was constructed by Invitrogen."

ORIGIN

Query Match 20.9%; Score 208; DB 14; Length 888;  
Best Local Similarity 100.0%; Pred. No. 1.2e-90;  
Matches 208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 631 AATAAATCTATATTTTGTGAAGGAGTGGTATTATATCTAGATTTCCAGTAGTTCTA 690  
DB 35 AATAAATCTATATTTTGTGAAGGAGTGGTATTATATCTAGATTTCCAGTAGTTCTA 94  
QY 691 AGTCTGTATTGTTTGTAAACAATGCGAGGTTTACACGCTCTATGCAATTGTACAAAA 750  
DB 95 AGTCTGTATTGTTTGTAAACAATGCGAGGTTTACACGCTCTATGCAATTGTACAAAA 154  
QY 751 AGTTATGAAGAACTACATGTAATCTTGTAGTCTAATAACTTCCGATTTCTTTATAT 810  
DB 155 AGTTATGAAGAACTACATGTAATCTTGTAGTCTAATAACTTCCGATTTCTTTATAT 214  
QY 811 GGAACGCATTTTGGGTTGTTTAAAAAATT 838  
DB 215 GGAACGCATTTTGGGTTGTTTAAAAAATT 242

RESULT 11  
AQ679243 525 bp DNA linear GSS 25-JUN-1999  
LOCUS  
DEFINITION HS 5417 A1 B06 T7A RPCI-11 Human Male BAC Library Homo sapiens  
genomic-clone Plate=993 Col=11 Row=C, genomic survey sequence.  
ACCESSION AQ679243  
VERSION AQ679243.1 GI:5228047  
KEYWORDS GSS.  
SOURCE  
ORGANISM Homo sapiens (human)  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 525)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.  
Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
Proc Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
93380589  
10449764  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: [jwallace@u.washington.edu](mailto:jwallace@u.washington.edu)  
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong ([pieter@dejong.med.buffalo.edu](mailto:pieter@dejong.med.buffalo.edu)). Clones may be purchased from BACPAC Resources ([http://bacpac.med.buffalo.edu/ordering\\_bac.htm](http://bacpac.med.buffalo.edu/ordering_bac.htm)) or from Research Genetics (<http://inforesgen.com>). BAC end Web Server: <http://www.htsc.washington.edu>  
Plate: 993 row: C column: 11  
Seq primer: T7  
Class: BAC ends  
High quality sequence stop: 525.  
Location/Qualifiers  
1. .525  
/organism="Homo sapiens"

FEATURES  
source







ORIGIN

Query Match 14.9%; Score 148; DB 28; Length 556;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-61;  
 Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	GTGGTTTGATTGATAGTAAATAAATAATGTTGTTTAAATACAAAGTACAGAGTAAGTAATCAAT	60
Db	153	GTGGTTTGATTGATAGTAAATAAATAATGTTGTTTAAATACAAAGTACAGAGTAAGTAATCAAT	94
Qy	61	CAATCACTCATAGCCAAAGTGGAAAAAGATGTATCCCATCATGGAATATTCCTGTTCIGAT	120
Db	93	CAATCACTCATAGCCAAAGTGGAAAAAGATGTATCCCATCATGGAATATTCCTGTTCIGAT	34
Qy	121	AGAAATCTTGCTTATCTATGGAATTC	148
Db	33	AGAAATCTTGCTTATCTATGGAATTC	6

Search completed: April 6, 2004, 17:40:09  
 Job time : 3674.8 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 13:02:17 ; Search time 4176.52 Seconds  
(without alignments)  
10336.278 Million cell updates/sec

Title: US-09-966-264D-2

Perfect score: 996

Sequence: 1 gtggttgattgatgtaa.....gtggttgatgcttaattaatt 996

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 3470272 seqs. 21671516995 residues

```

Word size : 0
Total number of hits satisfying chosen parameters: 6940544

```

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Listing first 45 summaries

Database : GenEmbl:\*

1: gb'ba:\*  
2: gb'htg:\*  
3: gb'in:\*  
4: gb'ov:\*  
5: gb'ov:\*  
6: gb'pat:\*  
7: gb'ph:\*  
8: gb'pl:\*  
9: gb'pr:\*  
10: gb'ro:\*  
11: gb'st:\*  
12: gb'sy:\*  
13: gb'un:\*  
14: gb'vi:\*  
15: em'ba:\*  
16: em'ba:\*  
17: em'hun:\*  
18: em'in:\*  
19: em'mu:\*  
20: em'om:\*  
21: em'or:\*  
22: em'ov:\*  
23: em'pat:\*  
24: em'ph:\*  
25: em'pl:\*  
26: em'ro:\*  
27: em'st:\*  
28: em'un:\*  
29: em'vi:\*  
30: em'htg' b  
31: em'htg' i  
32: em'htg' o  
33: em'htg' m  
34: em'htg' p  
35: em'htg' r  
36: em'htg' x  
37: em'htg' v  
38: em'sy:\*  
39: em'atgo  
40: em'atgo  
41: em'htgo

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
C	1	859	86.2	98056	9	AC006061 Homo sapi
	2	808	81.1	212120	9	AC023414 Homo sapi
	3	410	41.2	2463	6	BC187316 Sequence
	4	410	41.2	4658	9	BC028720 Homo sapi
	5	410	41.2	5339	6	AX538620 Sequence
	6	410	41.2	5417	6	AX538619 Sequence
	7	410	41.2	5462	6	AX538622 Sequence
	8	410	41.2	8689	6	AX538624 Sequence
	9	410	41.2	11443	6	AX538624 Sequence
	10	410	41.2	12057	6	AX538627 Sequence
	11	410	41.2	12446	9	HSDMDR Human mRNA
	12	410	41.2	12957	6	AX409637 Sequence
	13	410	41.2	13957	6	AX538581 Sequence
	14	410	41.2	13957	9	HUMDYS M18533 Homo sapien
	15	396	39.8	2148	11	GL5848 human STS C
	16	336	39.8	2691	6	AX538618 Sequence
	17	359	36.0	13977	6	AF220819 Sequence
	18	350	35.1	350	9	AF213444 Homo sapi
	19	298	29.0	2110	9	HUMDMDX M92650 Human Duche
	20	278	27.9	2563	9	AX129855 Homo sapi
	21	280	26.1	3163	6	E30223 Shortened d
	22	280	26.1	3172	6	E30222 Shortened d
	23	280	26.1	4075	6	E30221 Shortened d
	24	260	26.1	4402	6	E30219 Shortened d
	25	260	26.1	4402	6	E30220 Shortened d
	26	251	25.2	3747	6	E30218 Shortened d
	27	231	23.2	212120	2	AC023414 Homo sapi
	28	218	21.9	218	9	HUMDYST20 M86903 H. sapiens d
C	29	146	14.7	2148	11	GL5848 human STS C
	30	146	14.7	2563	9	AX129855 Homo sapi
	31	146	14.7	2691	6	AX538618 Sequence
	32	146	14.7	4658	9	BC028720 Homo sapi
	33	146	14.7	8689	6	AX538622 Sequence
	34	146	14.7	11443	6	AX538624 Sequence
	35	146	14.7	13957	6	AX409637 Sequence
	36	146	14.7	13957	6	AX538581 Sequence
	37	146	14.7	13957	9	HUMDYS M18533 Homo sapien
	38	146	14.7	98056	9	AC006061 Homo sapi
	39	145	14.6	2005	6	AX817318 Sequence
	40	126	12.7	13977	6	AX220819 Sequence
	41	102	10.2	13887	4	AF670485 Carls fam
	42	94	9.4	189131	10	AL645477 Mouse DNA
43	94	9.4	275578	2	AC108338 Rattus no	
C	44	94	9.4	279539	2	AC114184 Rattus no
	45	91	9.1	13815	6	AX306153 Sequence

## ALIGNMENTS

RESULT 1	AC006061/c	AC006061	98056 bp	DNA	linear	PRI 01-MAY-2002
LOCUS		Homo sapiens X BAC GSHB-19024			(Genome Systems Human BAC Library)	
DEFINITION		complete sequence.				
ACCESSION		AC006061				
VERSION		AC006061.1	GI:4204246			
KEYWORDS		HTG.				
SOURCE		Homo sapiens (human)				
ORGANISM		Homo sapiens				
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE		1 (bases 1 to 98056)				
AUTHORS		Muzny, D., Aronson, A.D., Bouck, J., Brundage, E., Bunac, C., Chen, Z., Niu, Ding, y. Dugan, S., Durbin, J., Forcum, J., Garcia, C., N. W.				

Correll, J.H., Gorrell, L.L., Hernandez, J., Jackson, L., Kondejewski, N., Leal, B., Lichtarge, O., Liu, W., Logan, O., Lu, J., Martinez, C., Oswal, G., Pampelli, L.R., Parish, B.J., Perez, L., Rashid, N.D., Rives, C., Scherer, S.E., Shen, H., Simon, M., Vo, Q., Williamson, A., Worley, K.C., Yu, W., Zhou, X., Nelson, D. and Gibbs, R.A.

Direct Submission  
Unpublished  
2 (bases 1 to 98056)  
Worley, K.C.

Direct Submission  
Submitted (26-NOV-1998) Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 98056)  
Worley, K.C.

Direct Submission  
Submitted (30-JAN-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
4 (bases 1 to 98056)  
Worley, K.C.

Direct Submission  
Submitted (02-FEB-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
5 (bases 1 to 98056)  
Worley, K.C.

Direct Submission  
Submitted (04-FEB-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
6 (bases 1 to 98056)  
Worley, K.C.

Direct Submission  
Submitted (28-MAR-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
7 (bases 1 to 98056)  
Worley, K.C.

Direct Submission  
Submitted (07-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
8 (bases 1 to 98056)  
Worley, K.C.

Direct Submission  
Submitted (01-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Jan 30, 1999 this sequence version replaced gi:4176317.  
INFORMATION: <http://gc.bcm.tmc.edu:8088/home.html> or email [gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

#### ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts. Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences. Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases.

FEATURES	Location/Qualifiers
source	1..98056
	/organism="Homo sapiens"
	/mol_type="Genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="X"
	/clone="GSHB-19024"
	1..1999
misc_feature	/note="Overlaps bases 151178..153177 of clone AC117405"
	/function="clone overlap"
repeat_region	40..82
	/rpt_family="L2"
repeat_region	378..507
	/rpt_family="L1MD3"
repeat_region	complement(518..570)
	/rpt_family="(GGA)n"
repeat_region	571..729
	/rpt_family="L1MC4"
repeat_region	730..987
	/rpt_family="MSTD"
repeat_region	complement(991..1400)
	/rpt_family="L2"
repeat_region	complement(1649..1779)
	/rpt_family="MIR"
repeat_region	complement(1783..2212)
	/rpt_family="LTR16A"
repeat_region	complement(3626..3656)
	/rpt_family="(GA)n"
repeat_region	complement(4713..5074)
	/rpt_family="THE1B"
repeat_region	5702..7108
	/rpt_family="L2"
repeat_region	complement(8670..8734)
	/rpt_family="MLT1G"
repeat_region	complement(9069..9144)
	/rpt_family="MLT1G"
repeat_region	complement(9254..9351)
	/rpt_family="MIR"
repeat_region	complement(9537..9688)
	/rpt_family="MER5A"
repeat_region	9797..9942
	/rpt_family="MIR"
repeat_region	complement(10403..10429)
	/rpt_family="AT rich"
repeat_region	complement(10673..10705)
	/rpt_family="(CA)n"
repeat_region	complement(11460..11755)
	/rpt_family="Alusx"
repeat_region	11977..12125
	/rpt_family="MER5A"
repeat_region	12556..13643
	/rpt_family="L1MD2"
repeat_region	13618..13808
	/rpt_family="L1MD3"
repeat_region	complement(13997..14297)
	/rpt_family="MER4D"
repeat_region	14327..14625
	/rpt_family="AluSg"
repeat_region	complement(14648..14915)
	/rpt_family="MER4D"
repeat_region	complement(14915..15147)
	/rpt_family="MER4D"
repeat_region	15286..15404
	/rpt_family="FLAM_C"





```

misc_feature 39491..43690
              /note="assembly_fragment"
misc_feature 43791..47621
              /note="assembly_fragment"
misc_feature 47722..53852
              /note="assembly_fragment"
              clone_end:T7
misc_feature 53953..63422
              /note="assembly_fragment"
misc_feature 63523..77361
              /note="assembly_fragment"
misc_feature 77462..92465
              /note="assembly_fragment"
misc_feature 92566..106951
              /note="assembly_fragment"
misc_feature 107052..129678

Query Match      81.1%; Score 808; DB 2; Length 212120;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 858; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTGGTTTCATGATAGTAAATGTTGTTTAAATCAAGTACAGAGTAAATCAAT 60
Db 180416 GTGGTTTCATGATAGTAAATGTTGTTTAAATCAAGTACAGAGTAAATCAAT 180475

QY 61 CAATCACTCATAGCAGAGTGGAAAGATGATCCATCATGGAATATTCCTGTTCTGAT 120
Db 180476 CAATCACTCATAGCAGAGTGGAAAGATGATCCATCATGGAATATTCCTGTTCTGAT 180535

QY 121 AGAATCTGTGCTATCTATGGAATCTTTTGATATATATTTACATTTGGAACCTCAAT 180
Db 180536 AGAATCTGTGCTATCTATGGAATCTTTTGATATATATTTACATTTGGAACCTCAAT 180595

QY 181 GTAGCTTGACATTTTCCATGTAACACACAGTAGCTGATCCACATTAAGCTGATACTA 240
Db 180596 GTAGCTTGACATTTTCCATGTAACACACAGTAGCTGATCCACATTAAGCTGATACTA 180655

QY 241 ACAACAACGTTAATGCTTCATTAATGAAGCTTTGCTTCTCTGGAACCTGGTGA 300
Db 180656 ACAACAACGTTAATGCTTCATTAATGAAGCTTTGCTTCTCTGGAACCTGGTGA 180715

QY 301 AATCAAACTGTTGTTGATACACCTCGATGAGCTTCTGTTGTTCTTCCACCAAGATG 360
Db 180716 AATCAAACTGTTGTTGATACACCTCGATGAGCTTCTGTTGTTCTTCCACCAAGATG 180775

QY 361 GGAATGATTTCCCAATGGCAAGAAACAGAGTATGCTATCTATCTGCACTTTTGTGA 420
Db 180776 GGAATGATTTCCCAATGGCAAGAAACAGAGTATGCTATCTATCTGCACTTTTGTGA 180835

QY 421 AAGTCTGTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 480
Db 180836 AAGTCTGTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 180895

QY 481 GCAGATGATTTGGGAGAGCGATGGAGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 540
Db 180896 GCAGATGATTTGGGAGAGCGATGGAGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 180955

QY 541 GCAGATGATTTGGGAGAGCGATGGAGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 600
Db 180956 GCAGATGATTTGGGAGAGCGATGGAGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 181015

QY 601 AAGAGATTTAGACAGTAAAGTGTACAGAAATAATCTATATTTTGTGAAGGCTAGTG 660
Db 181016 AAGAGATTTAGACAGTAAAGTGTACAGAAATAATCTATATTTTGTGAAGGCTAGTG 181075

QY 661 GTATTACTGTAGATTTCAAGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 720
Db 181076 GTATTACTGTAGATTTCAAGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 181135

QY 721 GTTTTACAGTCTATGCAATTTGACAAAAGTGTATAGAAAACCTACATGTAATAATCTTG 780
Db 181136 GTTTTACAGTCTATGCAATTTGACAAAAGTGTATAGAAAACCTACATGTAATAATCTTG 181195

```

```

QY 781 ATAGCTAAATAACTTGCATTTCTTTATATGAAAGCAATTTGGCTTGTAAAAAATTA 840
Db 181196 ATAGCTAAATAACTTGCATTTCTTTATATGAAAGCAATTTGGCTTGTAAAAAATTA 181255

QY 841 TAACAGTTTATAAGAAAGA 859
Db 181256 TAACAGTTTATAAGAAAGA 181274

RESULT 3
AX817316 2463 bp DNA linear PAT 10-DEC-2003
LOCUS Sequence 64 from Patent WO02081517.
DEFINITION AX817316
ACCESSION AX817316
VERSION AX817316.1 GI:397222703
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Decristofaro, M.F., Padigaru, M., Miller, C., Tchernev, V., Zhong, H.,
Zhong, M., Anderson, D., Ballinger, R., Gerlach, V., Spytek, K.A.,
Rastelli, L., Kekuda, R., Guo, X., Zerhusen, B., Andrew, D., Mezes, P.,
Patturajan, M., Burgess, C.E., Eisen, A., Wolenc, A., Baumgartner, J.,
Shimkets, R.A., Gusev, V., Vernec, C.A., Taupier, R.J., Pena, C.,
Shenoy, S., Li, L., Casman, S., Bolgog, F., Fernandes, E., Smithson, G.,
Malyankar, U., Tailon, B. and Liu, X.
TITLE Novel polypeptides and nucleic acids encoded thereby
JOURNAL Patent: WO 02081517-A 64 17-OCT-2002;
Curagen Corporation (US)
FEATURES
    source
        1..2463
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"
            <3..-2108
            /note="unnamed protein product"
            /codon_start=1
            /protein_id="CA884363.1"
            /db_xref="GI:397222704"
    CDS
        1..2463
            /translation="WASRHTPAPSGSAPSPAPRRLRLRSAPSCBLGRRRRRRRTFGE
            PGSGKLTPPLVTLDRGALAMREQLKHETQTTCDHPKMTLYQLSLADLNVRFS
            AYRTAKRLKOKALCWDLSLSACADLDONLAKNDQMDIOLINCLTITYDLRE
            OEHNLVNYPLCVDMLNLLNVDYDTGTRIRVLSFKTGIISLKAHLEDKYRLFK
            QVASTGFCDDRLGLLLHDSIQIPQLGEVAFSGSNLEPSVSCFOFANKKPEYA
            AFLDMRLPEQSMVPLVHRAAETAKHAKNCKECPILGFRVSLKHENDI
            COSCFPSRVAKGHKMHYPNVEYCTPTTSGEDVRDFAKVKNKFRKTRIFAKHPRNGY
            LPVQVLEGNMTEPTVLINFPVDSAPASSPOLSHDDTHSRHRYASLAEMNSNG
            SYLNDISPNESIDDEHLIIHQYCSLNQDSPLSQSPQALISLEBERGELERIL
            ADLEBNRLQAEYDRLKQOHEHKLSPSPPEMPTSPSPQPRDAELIAEAKLRQ
            KQLEARNMOLIDHNKQLESOLHRLRQLLEQQAQAKVNGTTPVSPSTSLQSDSSQP
            MLRVVGSQTSMSGEBDLSPQDTSGLSEVMQELNNSPPSSRHNVGSLFHMAAD
            LGHAMELSVSWTDEEGAE"
    ORIGIN
        Query Match      41.2%; Score 410; DB 6; Length 2463;
        Best Local Similarity 100.0%; Pred. No. 5.8e-203;
        Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 450 AGGACACAAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGAGAGCGATGAGTC 509
Db 2009 AGGACACAAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGAGAGCGATGAGTC 2068

QY 510 CTTAGTATCATGTCATGACAGATGAAGAAGAGCAGATAAATGTTTTTACAACCTCTCGATT 569
Db 2069 CTTAGTATCATGTCATGACAGATGAAGAAGAGCAGATAAATGTTTTTACAACCTCTCGATT 2128

QY 570 CCGCATGTTTATTAATATTCAACACAAAGAGATTAGACAGTAAGATTTTACAAG 629
Db 2129 CCGCATGTTTATTAATATTCAACACAAAGAGATTAGACAGTAAGATTTTACAAG 2188

```



QY 630 AAATAAATCTATATTTTGTGAAGGGTAGTGTATTACTAGTGTAGATTTTCAGTAGTTCT 689  
 Db 2189 AAATAAATCTATATTTTGTGAAGGGTAGTGTATTACTAGTGTAGATTTTCAGTAGTTCT 2248  
 QY 690 AAGTCTGTTATTTGTTTAAACAATGCGAGGTTTACACGTCATGCAATGTACAAAA 749  
 Db 2249 AAGTCTGTTATTTGTTTAAACAATGCGAGGTTTACACGTCATGCAATGTACAAAA 2308  
 QY 750 AAGTTATAAGAAAACACACATGTAATCTTGAATCTTAATAAATGTCATTTCTTTATA 809  
 Db 2309 AAGTTATAAGAAAACACACATGTAATCTTGAATCTTAATAAATGTCATTTCTTTATA 2368  
 QY 810 TCGAACGCAATTTGGGTTCTTTTAAAAATTTATACAGTTTATAAGAAAA 859  
 Db 2369 TCGAACGCAATTTGGGTTCTTTTAAAAATTTATACAGTTTATAAGAAAA 2418

RESULT 4  
 BC028720 4658 bp mRNA linear PRI 25-AUG-2003  
 LOCUS Homo sapiens dystrophin (muscular dystrophy, Duchenne and Becker  
 DEFINITION types) transcript variant Dp71b, mRNA (cDNA clone IMAGE:4822807),  
 complete cds.  
 ACCESSION BC028720.1 GI:20379675  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 4658)  
 Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zebberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E., Brownstein M.J., Udwin T.B., Toehiyuki S., Carninci P., Prange C., Raja S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S., Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E., Jones S.J. and Marra M.A.  
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 22388257  
 12477932  
 2 (bases 1 to 4658)  
 Strausberg R.  
 Direct Submission  
 Submitted (29-APR-2002) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki  
 Toehiyuki and Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbio.org>  
 contact: [amadanes@systemsbio.org](mailto:amadanes@systemsbio.org)  
 Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha  
 Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAC Plate: 46 Row: J Column: 3  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched rRNA gi: 5032296  
 This clone has the following problem: The cds is short compared to the longest cds in the locus.

FEATURES  
 Location/Qualifiers  
 1..4658  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4822807"  
 /tissue\_type="Testis"  
 /clone\_lib="NIH MGC\_97"  
 /lab\_host="DH10B"  
 /note="Vector: pBluescript"

gene  
 1..4658  
 /gene="DMD"  
 /note="synonyms: BMD, DXS142, DXS164, DXS206, DXS230, DXS239, DXS268, DXS269, DXS270, DXS272"  
 /db\_xref="LocusID:1756"  
 /db\_xref="MIM:300377"  
 130..2037  
 /codon\_start=1  
 /product="dystrophin, Dp71b isoform"  
 /protein\_id="AAH28720.1"  
 /db\_xref="GI:20379676"  
 /db\_xref="LocusID:1756"  
 /translation="MREQLKGHTTQTTWDPKTELYQSLADLNVRFSAYRTAMKLRLQKALCLDLLSLSAACDALDQNLKNDQMDILQIINCLTIDYRLQEHNLLVGVLCQVDMCNLLNVVDYTGTRIRVLSTKGIISLKAHLEKRYLFKQVASSYFCDQRLGLLHDSIQIIPROLGEVAFSGNSIVSRSCFQFANNKEIQAALPLDMWRLEFQSNWLPVLRHVAATAKQAQKCNICKPPIIGFRYSILKHFNYDICQSCFFSGRVAKGHQHPYMEYICTPTTSGEDVRDFAKVLKKNKRTKRYFAKHPRMGYLPQVTVLEGNMETPVTILINFVQSDASPSQLSHDDTHSRHRYASRLAEMNSGVSLNDSISPNESIDDEHLIIQYCSLNQSDSPSQSPAPQILISLESEEGELERILADLEENRNLQAEYDLKQOEHLGLSPPEMPMTSPQSPRDAELIAEKLRLQKFLVRLEARMQILEDHNKQLESOLHRLLOLLEPOQAQKVTGTVSSPSTLQSRSSQFMRLRVGSCDTSQSGEELSLPPQDTSTGLEVMQNLNNSPSSRGNVGSFLPHWADDLGRAMESLVSWNTDEGAE"

misc\_feature  
 844..981  
 /note="ZZ; Region: Zinc finger, ZZ type. Zinc finger present in dystrophin, CBP/p300. ZZ in dystrophin binds calmodulin Putative zinc finger"  
 /db\_xref="CDD:pfam00569"

ORIGIN  
 Query Match 41.2%; Score 410; DB 9; Length 4658;  
 Best Local Similarity 100.0%; Pred. No. 5.3e-203;  
 Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 450 AGGACACAATGTAGCAAGTCTTTTCCACATGGCAGATGATTGGGCAGAGCGATGGAGTC 509  
 Db 1938 AGGACACAATGTAGCAAGTCTTTTCCACATGGCAGATGATTGGGCAGAGCGATGGAGTC 1997  
 QY 510 CTTAGTATCATGTATGACAGATGAAGAGGACGAGATAAATGTTTACAACTCCTGATT 569  
 Db 1998 CTTAGTATCATGTATGACAGATGAAGAGGACGAGATAAATGTTTACAACTCCTGATT 2057  
 QY 570 CCCGATGTTTATATATATTTATATATTCATACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629  
 Db 2058 CCCGATGTTTATATATATTTATATATTCATACAAAGAGGATTAGACAGTAAGAGTTTACAAG 2117  
 QY 630 AAATAAATCTATATTTTGTGAAGGGTAGTGTATTACTAGTGTAGATTTTCAGTAGTTCT 689  
 Db 2118 AAATAAATCTATATTTTGTGAAGGGTAGTGTATTACTAGTGTAGATTTTCAGTAGTTCT 2177  
 QY 690 AAGTCTGTTATTTGTTTAAACAATGCGAGGTTTACACGTCATGCAATGTACAAAA 749  
 Db 2178 AAGTCTGTTATTTGTTTAAACAATGCGAGGTTTACACGTCATGCAATGTACAAAA 2237



Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 450 AGGACACAATGAGAAAGTCTTTTCCACATGCGCAGATGATTTGGCGCAGCGATGGAGTC 509  
DB 4614 AGGACACAATGAGAAAGTCTTTTCCACATGCGCAGATGATTTGGCGCAGCGATGGAGTC 4673  
QY 510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGATAAATGTTTACAACTCCTGATT 569  
DB 4674 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGATAAATGTTTACAACTCCTGATT 4733  
QY 570 CCCGATGTTTATTAATATTCATACAAAGAGGATAGACAGTAAGAGTTTACAAG 629  
DB 4734 CCCGATGTTTATTAATATTCATACAAAGAGGATAGACAGTAAGAGTTTACAAG 4793  
QY 630 AATTAATCTATATTTTGTGGAAGGTAGTGTATTTACTAGTATTCAGTAGTTTCT 689  
DB 4794 AATTAATCTATATTTTGTGGAAGGTAGTGTATTTACTGTAGATTTTCAGTAGTTTCT 4853  
QY 690 AAGTCTGTTATTTGTTTAAACAATGGCAGGTTTACACGTCATGAATTTGACAAAA 749  
DB 4854 AAGTCTGTTATTTGTTTAAACAATGGCAGGTTTACACGTCATGAATTTGACAAAA 4913  
QY 750 AAGTTATAAGAAACTACATGTAATCTTGTAGCTAAATAACTTGCATTTCTTTATA 809  
DB 4914 AAGTTATAAGAAACTACATGTAATCTTGTAGCTAAATAACTTGCATTTCTTTATA 4973  
QY 810 TGGAAACGCAATTTGGGTTGTTTAAATAATTATAACAGTTATAAAGAAAGA 859  
DB 4974 TGGAAACGCAATTTGGGTTGTTTAAATAATTATAACAGTTATAAAGAAAGA 5023

RESULT 8  
AX538622  
LOCUS AX538622 8689 bp DNA linear PAT 23-NOV-2002  
DEFINITION Sequence 42 from Patent WO0229056.  
ACCESSION AX538622  
VERSION AX538622.1 GI:25271171  
KEYWORDS synthetic construct  
SOURCE synthetic construct  
ORGANISM artificial sequences.  
REFERENCE 1  
AUTHORS Chamberlain, J.S. and Harper, S.O.  
TITLE Mini-dystrophin nucleic acid and peptide sequences  
JOURNAL Patent: WO 0229056-A 42 11-APR-2002;  
THE REGENTS OF THE UNIVERSITY OF MICHIGAN (US)  
FEATURES  
Location/Qualifiers  
source  
1..8689  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="Synthetic"

ORIGIN  
Query Match 41.2%; Score 410; DB 6; Length 8689;  
Best Local Similarity 100.0%; Pred. No. 4.9e-203;  
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 450 AGGACACAATGAGAAAGTCTTTTCCACATGCGCAGATGATTTGGCGCAGCGATGGAGTC 509  
DB 5985 AGGACACAATGAGAAAGTCTTTTCCACATGCGCAGATGATTTGGCGCAGCGATGGAGTC 6044  
QY 510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGATAAATGTTTACAACTCCTGATT 569  
DB 6045 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGATAAATGTTTACAACTCCTGATT 6104  
QY 570 CCCGATGTTTATTAATATTCATACAAAGAGGATAGACAGTAAGAGTTTACAAG 629  
DB 6105 CCCGATGTTTATTAATATTCATACAAAGAGGATAGACAGTAAGAGTTTACAAG 6164  
QY 630 AATTAATCTATATTTTGTGGAAGGTAGTGTATTTACTAGTATTTTCAGTAGTTTCT 689  
DB 6165 AATTAATCTATATTTTGTGGAAGGTAGTGTATTTACTAGTATTTTCAGTAGTTTCT 6224

QY 690 AAGTCTGTTATTTGTTTAAACAATGGCAGGTTTACACGTCATGAATTTGACAAAA 749  
DB 6225 AAGTCTGTTATTTGTTTAAACAATGGCAGGTTTACACGTCATGAATTTGACAAAA 6284  
QY 750 AAGTTATAAGAAACTACATGTAATCTTGTAGCTAAATAACTTGCATTTCTTTATA 809  
DB 6285 AAGTTATAAGAAACTACATGTAATCTTGTAGCTAAATAACTTGCATTTCTTTATA 6344  
QY 810 TGGAAACGCAATTTGGGTTGTTTAAATAATTATAACAGTTATAAAGAAAGA 859  
DB 6345 TGGAAACGCAATTTGGGTTGTTTAAATAATTATAACAGTTATAAAGAAAGA 6394

RESULT 9  
AX538624  
LOCUS AX538624 11443 bp DNA linear PAT 23-NOV-2002  
DEFINITION Sequence 44 from Patent WO0229056.  
ACCESSION AX538624  
VERSION AX538624.1 GI:25271175  
KEYWORDS synthetic construct  
SOURCE synthetic construct  
ORGANISM artificial sequences.  
REFERENCE 1  
AUTHORS Chamberlain, J.S. and Harper, S.O.  
TITLE Mini-dystrophin nucleic acid and peptide sequences  
JOURNAL Patent: WO 0229056-A 44 11-APR-2002;  
THE REGENTS OF THE UNIVERSITY OF MICHIGAN (US)  
FEATURES  
Location/Qualifiers  
source  
1..11443  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="Synthetic"

ORIGIN  
Query Match 41.2%; Score 410; DB 6; Length 11443;  
Best Local Similarity 100.0%; Pred. No. 4.7e-203;  
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 450 AGGACACAATGAGAAAGTCTTTTCCACATGCGCAGATGATTTGGCGCAGCGATGGAGTC 509  
DB 8739 AGGACACAATGAGAAAGTCTTTTCCACATGCGCAGATGATTTGGCGCAGCGATGGAGTC 8798  
QY 510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGATAAATGTTTACAACTCCTGATT 569  
DB 8799 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGATAAATGTTTACAACTCCTGATT 8858  
QY 570 CCCGATGTTTATTAATATTCATACAAAGAGGATAGACAGTAAGAGTTTACAAG 629  
DB 8859 CCCGATGTTTATTAATATTCATACAAAGAGGATAGACAGTAAGAGTTTACAAG 8918  
QY 630 AATTAATCTATATTTTGTGGAAGGTAGTGTATTTACTAGTATTTTCAGTAGTTTCT 689  
DB 8919 AATTAATCTATATTTTGTGGAAGGTAGTGTATTTACTAGTATTTTCAGTAGTTTCT 8978  
QY 690 AAGTCTGTTATTTGTTTAAACAATGGCAGGTTTACACGTCATGCAATTTGACAAAA 749  
DB 8979 AAGTCTGTTATTTGTTTAAACAATGGCAGGTTTACACGTCATGCAATTTGACAAAA 9038  
QY 750 AAGTTATAAGAAACTACATGTAATCTTGTAGCTAAATAACTTGCATTTCTTTATA 809  
DB 9039 AAGTTATAAGAAACTACATGTAATCTTGTAGCTAAATAACTTGCATTTCTTTATA 9098  
QY 810 TGGAAACGCAATTTGGGTTGTTTAAATAATTATAACAGTTATAAAGAAAGA 859  
DB 9099 TGGAAACGCAATTTGGGTTGTTTAAATAATTATAACAGTTATAAAGAAAGA 9148

RESULT 10  
AX538627  
LOCUS AX538627 12057 bp DNA linear PAT 23-NOV-2002

DEFINITION Sequence 47 from Patent WO0229056.  
ACCESSION AX538627  
VERSION AX538627.1 GI:25271181  
KEYWORDS synthetic construct  
SOURCE synthetic construct  
ORGANISM artificial sequences.  
REFERENCE 1  
AUTHORS Chamberlain, J.S. and Harper, S.O.  
TITLE Mini-dystrophin nucleic acid and peptide sequences  
JOURNAL Patent: WO 0229056-A 47 11-APR-2002;  
THE REGENTS OF THE UNIVERSITY OF MICHIGAN (US)  
FEATURES Location/Qualifiers  
source 1..12057  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="Synthetic"  
ORIGIN  
Query Match 41.2%; Score 410; DB 6; Length 12057;  
Best Local Similarity 100.0%; Pred. No. 4.7e-203;  
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 450 AGGACACATGATGAGAAAGTCTTTTCCACATGCGAGATGATTTGGCGAGCGATCGAGTC 509  
DB 11214 AGGACACATGATGAGAAAGTCTTTTCCACATGCGAGATGATTTGGCGAGCGATCGAGTC 11273  
QY 510 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGATAAATGTTTACAACTCCTGATT 569  
DB 11274 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGATAAATGTTTACAACTCCTGATT 11333  
QY 570 CCCGATCGTGTGTTTATAATATTTATACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629  
DB 11334 CCCGATCGTGTGTTTATAATATTTATACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11393  
QY 630 AAATAAATCTATPATTGTTGTAAGCGTAGTGATATTAATCTGATGATTTAGTAGTTTCT 689  
DB 11394 AAATAAATCTATPATTGTTGTAAGCGTAGTGATATTAATCTGATGATTTAGTAGTTTCT 11453  
QY 690 AAGTCGTGTTATGTTTGTAAATGTTTAAATGTTTAAATGTTTAAATGTTTAAATGTTTAA 749  
DB 11454 AAGTCGTGTTATGTTTGTAAATGTTTAAATGTTTAAATGTTTAAATGTTTAAATGTTTAA 11513  
QY 750 AAGTTTAAAGAAACTACATGTAATCTGATAGCTAAATGTAATCTGATAGCTAAATGTAATCTGATAGCT 809  
DB 11514 AAGTTTAAAGAAACTACATGTAATCTGATAGCTAAATGTAATCTGATAGCTAAATGTAATCTGATAGCT 11573  
QY 810 TGGACCGCATTTTGGGTTGTTTAAATGTTTAAATGTTTAAATGTTTAAATGTTTAAATGTTTAA 859  
DB 11574 TGGACCGCATTTTGGGTTGTTTAAATGTTTAAATGTTTAAATGTTTAAATGTTTAAATGTTTAA 11623  
RESULT 11  
HSDMDR 12446 bp mRNA linear PRI 12-SEP-1993  
LOCUS Human mRNA for dystrophin.  
DEFINITION X14298  
ACCESSION X14298.1 GI:30845  
VERSION Dmd gene; Duchenne muscular dystrophy; dystrophin.  
KEYWORDS Homo sapiens (human)  
SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 12446)  
AUTHORS Rosenthal, A., Speer, A., Billwiz, H., Cross, G.S., Forrest, S.M. and Davies, K.E.  
TITLE Two human cDNA molecules coding for the Duchenne muscular dystrophy (DMD) locus are highly homologous  
JOURNAL Nucleic Acids Res. 17 (13), 5391 (1989)  
MEDLINE 89345106  
PUBMED 2668885  
REFERENCE 2 (bases 1 to 12446)

Rosenthal, A.  
Direct Submission  
Submitted (09-FEB-1989) Rosenthal A., Akademie der Wissenschaften  
der DDR, Zentralinstitut fuer Molekularbiologie, Robert-Roessle  
Str.10, 1115 Berlin Buch, DDR  
see also M18533 and M20250 for Dmd seqs.: discrepancies compared to  
M18533 cDNA were located at x14298 pos. 496, 1772, 1965, 2449,  
3687, 4294, 4504, 5075, 5332, 5630 and 7194.  
Location/Qualifiers  
1..12446  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="X chromosome, Xp21."  
/tissue\_type="muscle"  
/dev\_stage="fetal and adult."  
99..11156  
/note="unnamed protein product; dystrophin (AA 1 - 3685)"  
/codon\_start=1  
/protein\_id="CAA32479.1"  
/db\_xref="GI:30846"  
/db\_xref="GOA:P11532"  
/translation="MLMVEEDCYEREDVQKFTTKVNAQFSKFGKHINLFSDL  
NQGRLLDLEGLTGQKLPKRGSTRVHALNNKALVLRNNVDLVNIGSTDIVDG  
NKKLTLLTWNILHWOVKNVKNMAGLOPTNSKILISWVRQSTRYPOVNVINFT  
TSMDGLALNALIHSHRPDLFDWNVVCOQATRLHAFNARYOLGTEKLLDPDQV  
DTYDPKSLIAYTSLFOVLPOQVSEIAIOEVMLEPRPKYKEEHFOLHQMHSYQ  
QITVSLAQGYRTISPKFRKSYATQAAYTTSPTSPSPPSQHLLEADKSGSSL  
MSEVNRIYQALBEVLSLWSEADTQAGEISNDVEVNDQVQTHHGHYMMDLTAH  
QVRGNILQGLKLTGKLTSEDETEVQWNNLNSWECRLQVQKQHLLEQLEQVRVN  
SLTNMNVVDESDDHATAALEQKLVLDGRWANTCRWTEDEWLLQDILKQWRLT  
DQTHMVALESEKEDAVNKTHHTGPKDNMELSSLOKLAVALKADLEKKGSMGLXSIK  
QDLSLTKNSYTKTEAWLNDNFARCDNMLVQLEKSTAQISOAVTTTQPSLTQTTVM  
EVTVITTEQLLVKHAQELPPPPKQKQITVSEIKELDLVDITELHSHITRSEA  
VQSPFAIFRKEGNSDLEKNAERERKFRKLQDASRSGQALVEQWNEGVNA  
DITQKASELNSRWIEFQCLLSERLNWYQNNIATFNQLOOLQEMTITNENLKIQ  
PTPSPFTKQKLTCDENVRNLGLQPIERLKIQSLAKEGQMFLLADFFVAF  
TNHFKVDFSDVQAKELQITFDLPMPRYQETMGARTWVQOSETKLSIPQLSVTDY  
EIMBOELGLOLQSLQOQSLVYLTSTVKEMKAPSEISRYQSEFEIEGRVK  
KLSOLVHEHCKLEBOMNKLKTONHIOQLTKVMAEDVLEKEPALGDSILKKOL  
KQRLVSDIQTIOPLNSWEGGQIKNEAPEPASPRLTELKSLTQMDHMCQVY  
ARKEAGGLEKTVSLQKOLSEHMTQABEELTDEPFEYKTPDELQAKFEMKRAK  
EQAQKQKALLESVNSVIAQAPVQAEKLELTTTNYQMLCTKQNKCKTLE  
EYVACWHELLSYLEKANKWNEVEFKTTENIPGAEEISEVLLSENLMSHSDNP  
NOTRIALQTLTGGVMDLEINEELTFNSRWELHEEAVRQKLESLQESQATENS  
LHLIOELTFIDKLAAYIADKVDAAQPOEAKIQSDLTSTHSEISEENKKNQKSEA  
AORVLSQIDVAKQKLDQVSMKFLQKQANFQRLQESKWLDEKVMHLPALETKSE  
QEVQSLAHCVNLYKLSSEKSEVMVLTGQIQVOKKOTENKPELDERVATLKLHY  
NELGAKVTERKQKLEKLSRMRKEMVLTENLARTDMLTKSASVEMSNLDE  
VAVGKATQKEIEKQVHLKSIITVEGALKVLGKETTLVDELTKSLNSNWIATVSRAE  
EWNLALLEYQKMETFDQVNDHTKWIIOADTLDESEKPKQKEDVLKRLKLEND  
IRPKVDSITRDQANLMANHDGCRKLVEPQISELNHRFAAI SHRIKTKGASIPKLELE  
QFNSDKLLEPLEAEIQGVNLKEDFNKMDNEQVGVKBLQDGLQRLQITDER  
KSEPIKIQOOLLTHNALDLSORRKALEISHOWYOYKQADLLKCLDIDIEKL  
TESKPAQRLNLAQIHTVRETMVMTEDMLEISYVSTYLTETHVTSQALLEVEQ  
ASIPERDERKIKEDIRELKKKEELNVAEROAGLSEDAAMAPEPTOIQLSKRWRE  
LUNAPDLKADFDLQKQESLNKISLQSSGRDITSHKTKAALPEVRL  
KEALSQLDFQKWKVNYQKQFDRSVKWRPHYDIQISKNTASILEQKTKQI  
PENWEHAKYTWLKELQDGIQGVTVRTNLGEBEIIQOQSKTASILEQKTKQI  
RWQEVKLSDRKRLKEQKILSEFQDLNFWLWLEADNIASTIPLEGKQQLKE  
KLEQVLLVEELPLROGILKNETGPGVLSAPISPEQDKENKLNQNLQVIVKS  
RALPEKQIEAQIKDLGLEKLELEQNLHLLMSPINQLEIVNQVQEGPFD  
VKETETAVQKQPDVEEILSKGHLKYEKATPQVYKRLDELSSEKAVNRLQELRA  
KQDPLAGVQTTIGASPTQVTLTQVTKETAISKLEMPSSLMLEVPALAFNRAT  
ELTMDLQDVIQKQVVMVGDLEINEMIIKQATMDLEQRPQGLSILEQAQNLK  
NKTNSQEARITIDIRIQNWDVEQHLNRRQQLNEMKDLQSLWLEAKAEQVL  
GOARAKLEWKEGPTVDVAIOKKTETKOLADLRQWNTVDVANDLAKLRLLYSAD  
DTEKVMITENASWRSIHKVSREAALEETHRLQFPDLDEKFLAWLAEATYAT  
NVLDQATRKERLLEDKSGVKELMKQWDLQGEIEATHDYVNLNENSKILSLSESD  
DAVLLQRRNNMNFKWSERKSLNRSLSERSSDQWKLHLSQLELLVWLQKDEL  
SRQAPITGGDFPAPQKQVHRAPKRELTKEPVIMSTLETVRLITLQTEPLEGKEDYQ

EPRELPEERAQNVRLRLKQAEVNTWEKLNHLSADWQKIDETLERLQELQEATD  
ETDLKQLQAEVYKSGWOPVGLDLSLDHLEKVKALRGAIAPLKNVSHVNDLARQL  
TLGIGLQSPYLNLSDELNTFWKLLQVAVEDRVQLHAEHDFGAPKNSHVLSTVQGP  
WERALSINKPYVYNHSTQTCWDPKMTLYOSLADLNVRPSAVETAKMLRLLOKA  
LCLDLNLSAACDALDQHNKLNKQNDPMDLIQINCLTIYDRLEQEHNNLYNPLVD  
MCLNLLNVDTGTRIRVLSFATGIIISLCAHLEDKYRILFKQVASSTGCDQRL  
GLLLHDSIQPRQGEVAFSGNSIEFVSFCFOFANNKPEIEAAUFLDMWRLEPQSM  
VWLPVJHRVAAAEAKHQAKNICKECPIIGFRYSRLKFNFDICOCFSFGRVAGXH  
KPHYPMVEYCPPTTSGEDVRDFAKVKNKFKTKYPAKHPMGVLPVQTVLEGNMET  
PVTILNFWFVDSAPASSPOLSHDDTHSRIEHYSRLAEMENSNGSYLNDISPNESID  
DEHLLIOHYCOSLANODPSLOPRSPAOLILSLESEERGELELILADLEENRNLOARY  
DRLKQEHKLSLPLPPEPMPTSPSPRDAELIAKLLRQHKGLERMLIEDH  
NKLESQHLRQLLEQPOAEKVGNTTVSSPSLSQSSDSQMLRUVVGSOTSISM  
GEEDLSPQDSTGTGLEWEMQLNNSPFSKRGNTPGKPMRDTM"

## ORIGIN

Query Match 41.2%; Score 410; DB 9; Length 12446;  
Best Local Similarity 100.0%; Pred. No. 4.7e-203;  
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 450 AGGACACAATGTAGGAAGTCCTTTCCACATGCGAGATGATTGGGCGAGCGATGGAGTC 509  
DB 11143 AGGACACAATGTAGGAAGTCCTTTCCACATGCGAGATGATTGGGCGAGCGATGGAGTC 11202

QY 510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGATAAATGTTTACAACTCCTGATT 569  
DB 11203 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGATAAATGTTTACAACTCCTGATT 11262

QY 570 CCCGCATGTTTTTATAATTTATACACAAGAGGATTAGACAGTAAGAGTTTACAAG 629  
DB 11263 CCCGCATGTTTTTATAATTTATACACAAGAGGATTAGACAGTAAGAGTTTACAAG 11322

QY 630 AAATAAATCTATATTTTGTGAAGGTAGTGGTATTATCTGTAGATTTCAGTAGTTCT 689  
DB 11323 AAATAAATCTATATTTTGTGAAGGTAGTGGTATTATCTGTAGATTTCAGTAGTTCT 11382

QY 690 AAGTCTGTTATTGTTTGTAAACAATGGCAGGTTTACACGTCATGCAATTTGACAAAA 749  
DB 11383 AAGTCTGTTATTGTTTGTAAACAATGGCAGGTTTACACGTCATGCAATTTGACAAAA 11442

QY 750 AAGTTATAGAAAACTACATGTAATAATCTTGATAGCTAAATAAATGCAATTTCTTTATA 809  
DB 11443 AAGTTATAGAAAACTACATGTAATAATCTTGATAGCTAAATAAATGCAATTTCTTTATA 11502

QY 810 TGGACGCAATTTGGTGTGTTTAAAAAATTATAACAGTTATAAAGAAAAA 859  
DB 11503 TGGACGCAATTTGGTGTGTTTAAAAAATTATAACAGTTATAAAGAAAAA 11552

RESULT 12  
AX409637  
LOCUS AX409637 13957 bp DNA linear PAT 14-JUN-2002  
DEFINITION Sequence 2284 from Patent WO0229103.  
ACCESSION AX409637  
VERSION AX409637.1 GI:21442342  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
REFERENCE 1  
AUTHORS Alvares,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.  
TITLE Gene expression profiles in liver cancer  
JOURNAL Patent: WO 0229103-A 2284 11-APR-2002;  
GENE LOGIC INC (US)  
FEATURES  
source  
1. .13957  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
/note="EMBL/GenBank Accession No. M18533"

## ORIGIN

Query Match 41.2%; Score 410; DB 6; Length 13957;  
Best Local Similarity 100.0%; Pred. No. 4.6e-203;  
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 450 AGGACACAATGTAGGAAGTCCTTTCCACATGCGAGATGATTGGGCGAGCGATGGAGTC 509  
DB 11253 AGGACACAATGTAGGAAGTCCTTTCCACATGCGAGATGATTGGGCGAGCGATGGAGTC 11312

QY 510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGATAAATGTTTACAACCTCCTGATT 569  
DB 11313 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGATAAATGTTTACAACCTCCTGATT 11372

QY 570 CCCGCATGTTTTTATAATTTATACACAAGAGGATTAGACAGTAAGAGTTTACAAG 629  
DB 11373 CCCGCATGTTTTTATAATTTATACACAAGAGGATTAGACAGTAAGAGTTTACAAG 11432

QY 630 AAATAAATCTATATTTTGTGAAGGTAGTGGTATTATCTGTAGATTTCAGTAGTTCT 689  
DB 11433 AAATAAATCTATATTTTGTGAAGGTAGTGGTATTATCTGTAGATTTCAGTAGTTCT 11492

QY 690 AAGTCTGTTATTGTTTGTAAACAATGGCAGGTTTACACGTCATGCAATTTGACAAAA 749  
DB 11493 AAGTCTGTTATTGTTTGTAAACAATGGCAGGTTTACACGTCATGCAATTTGACAAAA 11552

QY 750 AAGTTATAGAAAACTACATGTAATAATCTTGATAGCTAAATAAATGCAATTTCTTTATA 809  
DB 11553 AAGTTATAGAAAACTACATGTAATAATCTTGATAGCTAAATAAATGCAATTTCTTTATA 11612

QY 810 TGGACGCAATTTGGTGTGTTTAAAAAATTATAACAGTTATAAAGAAAAA 859  
DB 11613 TGGACGCAATTTGGTGTGTTTAAAAAATTATAACAGTTATAAAGAAAAA 11662

## RESULT 13

AX538581  
LOCUS AX538581 13957 bp DNA linear PAT 23-NOV-2002  
DEFINITION Sequence 1 from Patent WO0229056.  
ACCESSION AX538581  
VERSION AX538581.1 GI:25271086  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
REFERENCE 1  
AUTHORS Chamberlain,J.S. and Harper,S.C.  
TITLE Mini-dystrophin nucleic acid and peptide sequences  
JOURNAL Patent: WO 0229056-A 1 11-APR-2002;  
THE REGENTS OF THE UNIVERSITY OF MICHIGAN (US)  
FEATURES  
source  
1. .13957  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

## ORIGIN

Query Match 41.2%; Score 410; DB 6; Length 13957;  
Best Local Similarity 100.0%; Pred. No. 4.6e-203;  
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 450 AGGACACAATGTAGGAAGTCCTTTCCACATGCGAGATGATTGGGCGAGCGATGGAGTC 509  
DB 11253 AGGACACAATGTAGGAAGTCCTTTCCACATGCGAGATGATTGGGCGAGCGATGGAGTC 11312

QY 510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGATAAATGTTTACAACCTCCTGATT 569  
DB 11313 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGATAAATGTTTACAACCTCCTGATT 11372

QY 570 CCCGCATGTTTTTATAATTTATACACAAGAGGATTAGACAGTAAGAGTTTACAAG 629  
DB 11373 CCCGCATGTTTTTATAATTTATACACAAGAGGATTAGACAGTAAGAGTTTACAAG 11432

QY 630 AAATAAATCTATATTTTGTGAAGGTAGTGGTATTATCTGTAGATTTCAGTAGTTCT 689



Db 11253 AGGACACAACTAGGAAGTCTTTCCACATGGCAGATGATTTGGCAGAGCGATGGAGTC 11312

QY 510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGATAAATGTTTACAACTCCTGATT 569

Db 11313 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGATAAATGTTTACAACTCCTGATT 11372

QY 570 CCCGATCGTTTATATATATTTATACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629

Db 11373 CCCGATCGTTTATATATATTTATACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11432

QY 630 AAATAAATCTATATTTTGTGAAGGTTAGTGGTATTTATCTGTAGATTTCAAGTATTCT 689

Db 11433 AAATAAATCTATATTTTGTGAAGGTTAGTGGTATTTATCTGTAGATTTCAAGTATTCT 11492

QY 690 AAGTCGTGTTTGTGTTTACATGCGAGTTTACACCTCTATGCAATTTGACAAA 749

Db 11493 AAGTCGTGTTTGTGTTTACATGCGAGTTTACACCTCTATGCAATTTGACAAA 11552

QY 750 AAGTTTAAAGAACTACATGTAATCTTGATAGCTAAATAACTTGCATTTCTTTATA 809

Db 11553 AAGTTTAAAGAACTACATGTAATCTTGATAGCTAAATAACTTGCATTTCTTTATA 11612

QY 810 TGGACGCAATTTGGTGTGTTTAAATTTTAAACAGTTTATAAGAGAAAGA 859

Db 11613 TGGACGCAATTTGGTGTGTTTAAATTTTAAACAGTTTATAAGAGAAAGA 11662

RESULT 15

G15848 2148 bp DNA linear STS 19-JAN-1996

LOCUS human STS CHLC.UTR\_01924\_M18533.P56108 clone UTR\_01924\_M18533,

DEFINITION sequence tagged site.

ACCESSION G15848

VERSION G15848.1 GI:1161737

KEYWORDS STS; STS sequence; primer; sequence tagged site.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 2148)

AUTHORS Murray J., Sheffield, V., Weber, J.L., Duyk, G. and Buetow, K.H.

TITLE Cooperative Human Linkage Center

JOURNAL Unpublished (1995)

COMMENT Synonyms: UTR\_01924\_M18533, CHLC.UTR\_01924\_M18533.T36152

Contact: Dr. Jeffrey C. Murray

UofI

The University of Iowa

Department of Pediatrics, Iowa City, IA 52242, USA

Tel: (319) 356-3508

Fax: (319) 356-3347

Email: jeff-murray@uiowa.edu

Primer A: AACGATTTGGTGTGTTTA

Primer B: GATATACGCCAAAAGGATG

STS size: 189

PCR Profile:

denature: 30 seconds at 94 degrees C

annealing: 75 seconds at 55 degrees C

extension: 15 seconds at 72 degrees C

PCR cycles: 27

extension: 6 minutes at 72 degrees C

Protocol:

Template: 30ng genomic DNA

Primer: each 1.5 pmole

dNTPs: each 200 uM

Taq Polymerase: 0.3 units

Total Vol: 10 ul

Buffer:

MgCl2: 1.5mM

KCl: 50mM

Tris: 10mM

pH: 8.3

Prepared with primer pairs derived from M18533.

Location/Qualifiers

1. 2148

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

350..538

350..369

complement(519..538)

STS

primer\_bind

primer\_bind

ORIGIN

Query Match 39.8%; Score 396; DB 11; Length 2148;

Best Local Similarity 100.0%; Pred. No. 1.3e-195;

Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 464 GAAGTCTTTCCACATGGCAGATGATTTGGCAGAGCGATGAGTCCCTTAGTATCAGTCA 523

Db 1 GAAGTCTTTCCACATGGCAGATGATTTGGCAGAGCGATGAGTCCCTTAGTATCAGTCA 50

QY 524 TGACAGATGAAGAAGGAGCAGAAATAAATGTTTTTACAACTCCTGATTCGCGCATGTTTT 583

Db 61 TGACAGATGAAGAAGGAGCAGAAATAAATGTTTTTACAACTCCTGATTCGCGCATGTTTT 120

QY 584 ATATATTTATACACAAAGAGGATTAGACAGTAAGAGTTTACAAAGAAATTAATCTATAT 643

Db 121 ATATATTTATACACAAAGAGGATTAGACAGTAAGAGTTTACAAAGAAATTAATCTATAT 180

QY 644 TTTTGTGAAGGGTAGTGGTATTATCTGTAGATTTTCAAGTCTCTAGTCTGTTATTCT 703

Db 181 TTTTGTGAAGGGTAGTGGTATTATCTGTAGATTTTCAAGTCTCTAGTCTGTTATTCT 240

QY 704 TTTGTTAAACAATGGCAGGTTTTTACACGTCTATGCAATTTACAAAAAGTTATAGAAAA 763

Db 241 TTTGTTAAACAATGGCAGGTTTTTACACGTCTATGCAATTTACAAAAAGTTATAGAAAA 300

QY 764 CTACATGTAATAATCTGTAGATGCTAAATAACTTGCCTTTCTTTATATGGAACGCAATTTG 823

Db 301 CTACATGTAATAATCTGTAGATGCTAAATAACTTGCCTTTCTTTATATGGAACGCAATTTG 360

QY 824 GGTGTTTAAAAATTTATACAGTTTATAAGAAAGA 859

Db 361 GGTGTTTAAAAATTTATACAGTTTATAAGAAAGA 396

Search completed: April 6, 2004, 16:30:24

Job time : 4178.52 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 14:14:32 ; Search time 109.885 Seconds  
(without alignments)  
5030.075 Million cell updates/sec

Title: US-09-966-264D-2  
Perfect score: 996  
Sequence: 1 Gcgtttgatgtagtaaa.....gtgtttgatgtagtaatt 996

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/Backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	359	36.0	13977	4	US-09-484-970B-60
2	126	12.7	13977	4	US-09-484-970B-60
3	91	9.1	19307	3	US-08-836-022A-10
4	91	9.1	19307	3	US-09-427-048A-10
5	42	4.2	19307	3	US-08-836-022A-10
6	42	4.2	19307	3	US-09-427-048A-10
7	23	2.3	434	3	US-08-943-731-127
8	23	2.3	24183	3	US-08-943-731-3
9	20	2.0	891	4	US-08-936-165A-90
10	20	2.0	1201	3	US-09-274-642-1
11	19	1.9	4766	5	PCT-US93-07261-10
12	19	1.9	9139	4	US-09-322-478-22
13	18	1.8	25	2	US-08-676-782-4
14	18	1.8	342	1	US-08-676-782-11
15	18	1.8	393	1	US-08-248-505-1
16	18	1.8	393	2	US-08-676-782-6
17	18	1.8	750	4	US-09-107-532A-866
18	18	1.8	889	2	US-08-935-886-7
19	18	1.8	889	2	US-08-935-886-13
20	18	1.8	1098	4	US-08-956-171E-253
21	18	1.8	1349	2	US-08-676-782-10
22	18	1.8	1824	4	US-09-489-039A-4969
23	18	1.8	1819	4	US-09-398-179-1
24	18	1.8	1819	4	US-09-398-179-2
25	18	1.8	1966	4	US-09-221-017B-964
26	18	1.8	2076	4	US-09-134-001C-1838
27	18	1.8	2382	4	US-09-107-532A-2569

ALIGNMENTS

RESULT 1  
US-09-484-970B-60  
; Sequence 60, Application US/09484970B  
; Patent No. 6426186  
; GENERAL INFORMATION:  
; APPLICANT: Jones, Karen A.  
; APPLICANT: Volkmut, Wayne  
; APPLICANT: Walker, Michael G.  
; TITLE OF INVENTION: BONE REMODELING GENES  
; FILE REFERENCE: PB-0014 US  
; CURRENT APPLICATION NUMBER: US/09/484,970B  
; CURRENT FILING DATE: 2000-01-18  
; NUMBER OF SEQ ID NOS: 172  
; SOFTWARE: PERL Program  
; SEQ ID NO 60  
; LENGTH: 13977  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. 6426186 229357.11CBI  
; NAME/KEY: unsure  
; LOCATION: 11721-11761, 12294, 13969  
; OTHER INFORMATION: a, t, c, g, or other  
US-09-484-970B-60

Query Match	36.0%	Score 359;	DB 4;	Length 13977;
Best Local Similarity	99.8%	Pred. No. 6.3e-170;		
Matches	409;	Conservative	0;	Mismatches 1;
Indels	0;	Gaps	0;	
Qy	450	AGGACCAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGATC	509	
Db	11254	AGGACCAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGATC	11313	
Qy	510	CTTAGTATCATGATGACAGATGACAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAGC	569	
Db	11314	CTTAGTATCATGATGACAGATGACAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAGC	11373	
Qy	570	CCGCGATGGTTTATTAATATTCATACACAAAGAGGATTAGACAGTAAAGAGTTTACAAG	629	
Db	11374	CCGCGATGGTTTATTAATATTCATACACAAAGAGGATTAGACAGTAAAGAGTTTACAAG	11433	
Qy	630	AAATAATCTATATTTTGTGAAGGTAGTGGTATTATCTAGTATGATTTAGTATTTCT	689	
Db	11434	AAATAATCTATATTTTGTGAAGGTAGTGGTATTATCTAGTATGATTTAGTATTTCT	11493	
Qy	690	AAGTCTGTTATCTTTTGTAAACAAAGTGGAGGTTTACAGTCTATGCAATTTCAAAA	749	
Db	11494	AAGTCTGTTATCTTTTGTAAACAAAGTGGAGGTTTACAGTCTATGCAATTTCAAAA	11553	



QY 750 AAGTTATAGAAACTACATGTAATAATCTGTAGTAACTAAATAAATGCAATTCCTTTATA 809  
DB 11554 AAGTTATAGAAACTACATGTAATAATCTGTAGTAACTAAATAAATGCAATTCCTTTATA 11613  
QY 810 TGGACGCAATTTGGTGTGTTTAAATAATTATAACAGTTATAAAGAAAGA 859  
DB 11614 TGGACGCAATTTGGTGTGTTTAAATAATTATAACAGTTATAAAGAAAGA 11663

## RESULT 2

US-09-484-970B-60/c  
; Sequence 60, Application US/09484970B  
; Patent No. 6426186  
; GENERAL INFORMATION:  
; APPLICANT: Jones, Karen A.  
; APPLICANT: Volkmut, Wayne  
; APPLICANT: Walker, Michael G.  
; TITLE OF INVENTION: BONE REMODELING GENES  
; FILE REFERENCE: PB-0014 US  
; CURRENT APPLICATION NUMBER: US/09/484,970B  
; CURRENT FILING DATE: 2000-01-18  
; NUMBER OF SEQ ID NOS: 172  
; SOFTWARE: PERL Program  
; SEQ ID NO 60  
; LENGTH: 13977  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc.feature  
; OTHER INFORMATION: Incyte ID No. 6426186 229357.11CB1  
; NAME/KEY: unsure  
; LOCATION: 11721-11761, 12294, 13969  
; OTHER INFORMATION: a, t, c, g, or other  
US-09-484-970B-60

Query Match 12.7%; Score 126; DB 4; Length 13977;

Best Local Similarity 100.0%; Pred. No. 2.5e-53;  
Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 851 AAAGAAGAAATTATAAGGAAAGAAAGAAATAACGCAATGCAAGTGGTGAAGCTGTGAA 910  
DB 13315 AAAGAAGAAATTATAAGGAAAGAAAGAAATAACGCAATGCAAGTGGTGAAGCTGTGAA 13256  
QY 911 CTCAGGTGTGCACAAATATCAGGAACACCCCAACCAAGTGAAGTGAATACATGA 970  
DB 13255 CTCAGGTGTGCACAAATATCAGGAACACCCCAACCAAGTGAAGTGAATACATGA 13196  
QY 971 GAAGCC 976  
DB 13195 GAAGCC 13190

## RESULT 3

US-08-836-022A-10/c  
; Sequence 10, Application US/08836022A  
; Patent No. 6001557  
; GENERAL INFORMATION:  
; APPLICANT: Trustees of the University of Pennsylvania  
; APPLICANT: Wilson, James M.  
; APPLICANT: Fisher, Krishna J.  
; APPLICANT: Chen, Shu-Jen  
; APPLICANT: Weitzman, Matthew  
; TITLE OF INVENTION: Improved Adenovirus Virus and  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Howson and Howson  
; STREET: Spring House Corporate Cntr, P O Box 457  
; City: Spring House  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19477  
; COMPUTER READABLE FORM: disk  
; MEDIUM TYPE: Floppy

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/836,022A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA: US 08/331,381  
FILING DATE: 28-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: GNVEN.008PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-540-9200  
TELEFAX: 215-540-5818  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19307 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: cdna  
US-08-836-022A-10

Query Match 9.1%; Score 91; DB 3; Length 19307;

Best Local Similarity 100.0%; Pred. No. 8.4e-36;  
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 665 TATACGTAGATTCTAGTAGTTCTTAAGTCTGTATTTGTTTAACTGCGAGTTT 724  
DB 3225 TATACGTAGATTCTAGTAGTTCTTAAGTCTGTATTTGTTTAACTGCGAGTTT 3166  
QY 725 TACAGTCTATGCAATTGTACAAAAAAGTTA 755  
DB 3165 TACAGTCTATGCAATTGTACAAAAAAGTTA 3135

## RESULT 4

US-09-427-048A-10/c  
; Sequence 10, Application US/09427048A  
; Patent No. 6203975  
; GENERAL INFORMATION:  
; APPLICANT: Trustees of the University of Pennsylvania  
; APPLICANT: Wilson, James M.  
; APPLICANT: Fisher, Krishna J.  
; APPLICANT: Chen, Shu-Jen  
; APPLICANT: Weitzman, Matthew  
; TITLE OF INVENTION: Improved Adenovirus Virus and  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Howson and Howson  
; STREET: Spring House Corporate Cntr, P O Box 457  
; City: Spring House  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19477  
; COMPUTER READABLE FORM: floppy disk  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/427,048A  
; FILING DATE: 21-Oct-1999  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/836,022  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bak, Mary E.

```

US-08-836-022A-10
Query Match      4.2%; Score 42; DB 3; Length 19307;
Best Local Similarity 100.0%; Pred. No. 2.8e-11;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 934 AACACCCCCAAAACCAAGTGAGGTAGAGTAAGATAGCATGAGAGC 975
    |||||||
Db 1563 AACACCCCCAAAACCAAGTGAGGTAGAGTAAGATAGCATGAGAGC 1604
    |||||||

RESULT 6
US-09-427-048A-10
; Sequence 10, Application US/09427048A
; Patent No. 6203975
; GENERAL INFORMATION:
; APPLICANT: Trustees of the University of Pennsylvania
; Wilson, James M.
; Fisher, Krishna J.
; Chen, Shu-Jen
; Weitzman, Matthew
; TITLE OF INVENTION: Improved Adenovirus Virus and
; Methods of Use Thereof
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P O Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/427,048A
; FILING DATE: 21-Oct-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,022
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: GNVPN.008PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19307 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: CDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-427-048A-10

Query Match      4.2%; Score 42; DB 3; Length 19307;
Best Local Similarity 100.0%; Pred. No. 2.8e-11;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 934 AACACCCCCAAAACCAAGTGAGGTAGAGTAAGATAGCATGAGAGC 975
    |||||||
Db 1563 AACACCCCCAAAACCAAGTGAGGTAGAGTAAGATAGCATGAGAGC 1604
    |||||||

RESULT 7
US-08-943-731-127
; Sequence 127, Application US/08943731
; Patent No. 6265157

```

```

; GENERAL INFORMATION:
; APPLICANT: PROCKOP, DARWIN J.
; APPLICANT: SPOTILA, LORETTA D.
; APPLICANT: DELTAS, CONSTANTINOS D.
; APPLICANT: SEREDA, LARISA
; APPLICANT: LARSON, ANDREA W.
; APPLICANT: PACK, MICHAEL
; APPLICANT: COLIGE, ALAIN
; APPLICANT: EARLY, JAMES
; APPLICANT: KORKKO, JARMO
; APPLICANT: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; APPLICANT: ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND
; STREET: FLR.
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-7086
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/943,731
; FILING DATE: 03-OCT-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/212,322
; FILING DATE: 14-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/803,628
; FILING DATE: 03-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: DOYLE LEARY Ph.D., KATHRYN
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: 9598-27
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-965-1284
; TELEFAX: 215-567-2991
; TELEX: 831-494
; INFORMATION FOR SEQ ID NO: 127:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 434 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-943-731-127

Query Match 2.3%; Score 23; DB 3; Length 434;
Best Local Similarity 100.0%; Pred. No. 0.084;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 423 GTCTGCTCTTCTCTCTTGT 445
Db 395 GTCTGCTCTTCTCTCTTGT 417

RESULT 8
US-08-943-731-3
; Sequence 3, Application US/08943731
; Patent No. 6265157
; GENERAL INFORMATION:
; APPLICANT: PROCKOP, DARWIN J.
; APPLICANT: SPOTILA, LORETTA D.
; APPLICANT: DELTAS, CONSTANTINOS D.
; APPLICANT: SEREDA, LARISA
; APPLICANT: LARSON, ANDREA W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
; TITLE OF INVENTION: ALTERED TYPE I OR TYPE IX COLLAGEN GENE SEQUENCES
; NUMBER OF SEQUENCES: 666
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; STREET: ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND
; STREET: FLR.
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-7086
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/943,731
; FILING DATE: 03-OCT-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/212,322
; FILING DATE: 14-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/803,628
; FILING DATE: 03-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: DOYLE LEARY Ph.D., KATHRYN
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: 9598-27
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-965-1284
; TELEFAX: 215-567-2991
; TELEX: 831-494
; INFORMATION FOR SEQ ID NO: 127:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 434 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-943-731-127

Query Match 2.3%; Score 23; DB 3; Length 434;
Best Local Similarity 100.0%; Pred. No. 0.084;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 423 GTCTGCTCTTCTCTCTTGT 445
Db 395 GTCTGCTCTTCTCTCTTGT 417

RESULT 9
US-08-936-165A-90
; Sequence 90, Application US/08936165A
; Patent No. 6348582
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Burdham, Martin
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Lonetto, Michael
; APPLICANT: Nicholas, Richard
; APPLICANT: Pratt, Julie
; APPLICANT: Reichard, Richard
; APPLICANT: Rosenberg, Martin
; APPLICANT: Ward, Judith
; TITLE OF INVENTION: NO. 6348582el Prokaryotic Polynucleotides,

```

```

; APPLICANT: PACK, MICHAEL
; APPLICANT: COLIGE, ALAIN
; APPLICANT: EARLY, JAMES
; APPLICANT: KORKKO, JARMO
; APPLICANT: ALA-KOKKO, LEENA, et al.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
; TITLE OF INVENTION: ALTERED TYPE I OR TYPE IX COLLAGEN GENE SEQUENCES
; NUMBER OF SEQUENCES: 666
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; STREET: ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND
; STREET: FLR.
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-7086
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/943,731
; FILING DATE: 03-OCT-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/212,322
; FILING DATE: 14-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/803,628
; FILING DATE: 03-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: DOYLE LEARY Ph.D., KATHRYN
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: 9598-27
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-965-1284
; TELEFAX: 215-567-2991
; TELEX: 831-494
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 24183 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-943-731-3

Query Match 2.3%; Score 23; DB 3; Length 24183;
Best Local Similarity 100.0%; Pred. No. 0.09;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 423 GTCTGCTCTTCTCTCTTGT 445
Db 10136 GTCTGCTCTTCTCTCTTGT 10158

RESULT 9
US-08-936-165A-90
; Sequence 90, Application US/08936165A
; Patent No. 6348582
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Burdham, Martin
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Lonetto, Michael
; APPLICANT: Nicholas, Richard
; APPLICANT: Pratt, Julie
; APPLICANT: Reichard, Richard
; APPLICANT: Rosenberg, Martin
; APPLICANT: Ward, Judith
; TITLE OF INVENTION: NO. 6348582el Prokaryotic Polynucleotides,

```

## TITLE OF INVENTION: Polypeptides and Their Uses

NUMBER OF SEQUENCES: 534  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-0939

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASECQ for Windows Version 2.0

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/936,165A

## FILING DATE: 24-SEP-1997

## CLASSIFICATION: 536

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/027,032

## FILING DATE: 24-SEP-1996

## ATTORNEY/AGENT INFORMATION:

NAME: Gimmi, Edward R

## REGISTRATION NUMBER: 38,891

REFERENCE/DOCKET NUMBER: P50549

## TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-270-4478

TELEFAX: 610-270-5090

## TELEX:

## INFORMATION FOR SEQ ID NO: 90:

## SEQUENCE CHARACTERISTICS:

LENGTH: 891 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: Genomic DNA

US-08-936-165A-90

Query Match 2.0%; Score 20; DB 4; Length 891;

Best Local Similarity 100.0%; Pred. No. 2.7;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 686 TTCTAAGTCGTGTTATTGTTT 705

Db 579 TTCTAAGTCGTGTTATTGTTT 598

## RESULT 10

US-09-274-642-1/c

Sequence 1, Application US/09274642A

Patent No. 6071729

## GENERAL INFORMATION:

APPLICANT: Jeffries, Thomas W.

APPLICANT: Shi, Nian-Qing

TITLE OF INVENTION: Disruption of cytochrome c gene in xylose-fermenting yeast

FILE REFERENCE: cytochrome c

CURRENT APPLICATION NUMBER: US/09/274,642A

CURRENT FILING DATE: 1999-03-23

EARLIER APPLICATION NUMBER: 60/080,493

EARLIER FILING DATE: 1998-04-02

NUMBER OF SEQ ID NOS: 42

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 1

LENGTH: 1201

TYPE: DNA

ORGANISM: Pichia stipitis

## FEATURE:

NAME/KEY: CDS

LOCATION: (610)..(942)

US-09-274-642-1

## Query Match

Best Local Similarity 2.0%; Score 20; DB 3; Length 1201;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 623 TTACAAGAAATAAATCTATA 642

Db 1062 TTACAAGAAATAAATCTATA 1043

## RESULT 11

PCT-US93-07261-10

Sequence 10, Application PC/TUS9307261

## GENERAL INFORMATION:

TITLE OF INVENTION: PfEMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:

ADDRESSEE: John H. C. Blasdale

STREET: One Giralda Farms

CITY: Madison

STATE: New Jersey

COUNTRY: USA

ZIP: 07940-1000

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: Apple Macintosh

OPERATING SYSTEM: Macintosh 6.0.5

SOFTWARE: Microsoft Word 5.1a

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US93/07261

FILING DATE: 19930805

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/927,531

FILING DATE: 07-AUG-1992

ATTORNEY/AGENT INFORMATION:

NAME: Blasdale, John H. C.

REGISTRATION NUMBER: 31,895

REFERENCE/DOCKET NUMBER: DX0288K

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-822-7398

TELEFAX: 201-822-7039

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 4766 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA to mRNA

ORGANISM: Plasmodium falciparum

STRAIN: Malayan Camp

IMMEDIATE SOURCE:

CLONE: p2b1p12-1

FEATURE:

NAME/KEY: CDS

LOCATION: 3..4766

PCT-US93-07261-10

Query Match 1.9%; Score 19; DB 5; Length 4766;

Best Local Similarity 100.0%; Pred. No. 8.8;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 864 TAAAGGAAAAAGAAATAA 882

Db 190 TAAAGGAAAAAGAAATAA 208

## RESULT 12

US-09-322-478-22/c

Sequence 22, Application US/09322478

Patent No. 6331662

## GENERAL INFORMATION:

APPLICANT: Wright, David A.

APPLICANT: Voytas, Daniel F.

TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto

FILE REFERENCE: P-1065 ISURF Plant Retroelement

;; CURRENT APPLICATION NUMBER: US/09/322,478  
;; CURRENT FILING DATE: 1999-05-28  
;; EARLIER APPLICATION NUMBER: 60/087125  
;; EARLIER FILING DATE: 1998-05-29  
;; NUMBER OF SEQ ID NOS: 41  
;; SOFTWARE: Patent in Ver. 2.0  
;; SEQ ID NO 22  
;; LENGTH: 9139  
;; TYPE: DNA  
;; ORGANISM: Glycine max  
US-09-322-478-22

Query Match 1.8%; Score 19; DB 4; Length 9139;  
Best Local Similarity 100.0%; Pred. No. 8.9;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 528 AGATGAAGAAGGAGCAGAA 546  
|||||  
DB 5801 AGATGAAGAAGGAGCAGAA 5783

RESULT 13  
US-08-676-782-4/c  
;; Sequence 4, Application US/08676782  
;; Patent No. 5976732  
;; GENERAL INFORMATION:  
;; APPLICANT: CHEUNG, Ambrose  
;; APPLICANT: FISCHETTI, Vincent A.  
;; TITLE OF INVENTION: REGULATION OF EXOPROTEIN IN  
;; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS  
;; NUMBER OF SEQUENCES: 18  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.  
;; STREET: P.O. Box 1404  
;; CITY: Alexandria  
;; STATE: Virginia  
;; COUNTRY: United States  
;; ZIP: 22313-1404  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent in Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/676,782  
;; FILING DATE: 08-JUL-1996  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/248,505  
;; FILING DATE: 25-MAY-1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: McGowan, Malcolm K.  
;; REGISTRATION NUMBER: 39,300  
;; REFERENCE/DOCKET NUMBER: 016921-092  
;; TELEPHONE: (703) 836-6620  
;; TELEFAX: (703) 836-2021  
;; INFORMATION FOR SEQ ID NO: 4:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 25 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
US-08-676-782-4

Query Match 1.8%; Score 18; DB 2; Length 25;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 431 TTCTTCTCTTTGTTTC 448  
|||||  
DB 24 TTCTTCTCTTTGTTTC 7

RESULT 14  
US-08-676-782-11/c  
;; Sequence 11, Application US/08676782  
;; Patent No. 5976732  
;; GENERAL INFORMATION:  
;; APPLICANT: CHEUNG, Ambrose  
;; APPLICANT: FISCHETTI, Vincent A.  
;; TITLE OF INVENTION: REGULATION OF EXOPROTEIN IN  
;; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS  
;; NUMBER OF SEQUENCES: 18  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.  
;; STREET: P.O. Box 1404  
;; CITY: Alexandria  
;; STATE: Virginia  
;; COUNTRY: United States  
;; ZIP: 22313-1404  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent in Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/676,782  
;; FILING DATE: 08-JUL-1996  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/248,505  
;; FILING DATE: 25-MAY-1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: McGowan, Malcolm K.  
;; REGISTRATION NUMBER: 39,300  
;; REFERENCE/DOCKET NUMBER: 016921-092  
;; TELEPHONE: (703) 836-6620  
;; TELEFAX: (703) 836-2021  
;; INFORMATION FOR SEQ ID NO: 11:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 342 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; FEATURE:  
;; NAME/KEY: misc\_feature  
;; LOCATION: 157  
;; OTHER INFORMATION: /note= "Nucleotide at position 157  
;; OTHER INFORMATION: is N wherein N = C or T."  
;; FEATURE:  
;; NAME/KEY: misc\_feature  
;; LOCATION: 340  
;; OTHER INFORMATION: /note= "Nucleotide at position 340  
;; OTHER INFORMATION: is N wherein N = C or T."  
US-08-676-782-11

Query Match 1.8%; Score 18; DB 2; Length 342;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 431 TTCTTCTCTTTGTTTC 448  
|||||  
DB 150 TTCTTCTCTTTGTTTC 133

RESULT 15  
US-08-248-505-1/c  
;; Sequence 1, Application US/08248505  
;; Patent No. 5587288  
;; GENERAL INFORMATION:  
;; APPLICANT: CHEUNG, Ambrose  
;; APPLICANT: FISCHETTI, Vincent A.

```

; TITLE OF INVENTION: REGULATION OF EXOPROTEIN IN
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/248,505
; FILING DATE: 24-MAY-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Crane-Feury, Sharon E
; REGISTRATION NUMBER: 36,113
; REFERENCE/DOCKET NUMBER: 016921-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 393 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 19..390
; US-08-248-505-1

```

```

Query Match      1.8%; Score 18; DB 1; Length 393;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      431 TTCTTTCTCTTTGTTTC 448
          |||||
Db      168 TTCTTTCTCTTTGTTTC 151

```

Search completed: April 6, 2004, 17:42:28  
Job time : 111.385 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 15:01:53 ; Search time 449.211 Seconds  
(without alignments)  
8309.593 Million cell updates/sec

Title: US-09-966-264D-2

Perfect score: 996  
Sequence: 1 ggggttgattgatagataaa.....gtgtttgatgtaataatt 996

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2470632 seqs, 1873875610 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4941264

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications NA:\*\*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US05\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
16: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*  
18: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	996	100.0	996	US-09-966-264-2	Sequence 2, Appli
2	410	41.2	5339	US-10-149-736-40	Sequence 40, Appli
3	410	41.2	5417	US-10-149-736-39	Sequence 39, Appli
4	410	41.2	5462	US-10-149-736-41	Sequence 41, Appli
5	410	41.2	8689	US-10-149-736-42	Sequence 42, Appli
6	410	41.2	14443	US-10-149-736-44	Sequence 44, Appli
7	410	41.2	12057	US-10-149-736-47	Sequence 47, Appli
8	410	41.2	13957	US-09-782-378A-22	Sequence 22, Appli
9	410	41.2	13957	US-09-880-107-2284	Sequence 2284, Ap
10	410	41.2	13957	US-10-149-736-1	Sequence 1, Appli
11	410	41.2	14069	US-10-342-887-434	Sequence 434, App
12	410	41.2	14082	US-10-342-887-981	Sequence 981, App
13	410	41.2	14082	US-10-341-434-108	Sequence 108, App
14	396	39.8	2691	US-10-149-736-38	Sequence 38, Appli
15	146	14.7	430	US-09-796-692-3505	Sequence 3505, Ap

c 16	146	14.7	430	14	US-10-040-862-3505	Sequence 3505, Ap
c 17	146	14.7	430	15	US-10-057-475B-3505	Sequence 3505, Ap
c 18	146	14.7	430	15	US-10-154-884B-3505	Sequence 3505, Ap
c 19	146	14.7	2691	15	US-10-149-736-38	Sequence 38, Appli
c 20	146	14.7	8689	15	US-10-149-736-42	Sequence 42, Appli
c 21	146	14.7	11443	15	US-10-149-736-44	Sequence 44, Appli
c 22	146	14.7	13957	9	US-09-782-378A-22	Sequence 22, Appli
c 23	146	14.7	13957	9	US-09-880-107-2284	Sequence 2284, Ap
c 24	146	14.7	13957	15	US-10-149-736-1	Sequence 1, Appli
c 25	146	14.7	14069	12	US-10-342-887-434	Sequence 434, App
c 26	146	14.7	14082	12	US-10-342-887-981	Sequence 981, App
c 27	146	14.7	14082	15	US-10-341-434-108	Sequence 108, App
c 28	137	13.8	137	9	US-09-966-264-1	Sequence 1, Appli
c 29	91	9.1	13815	15	US-10-149-736-2	Sequence 2, Appli
c 30	42	4.2	42	9	US-09-966-264-8	Sequence 8, Appli
c 31	42	4.2	13815	15	US-10-149-736-2	Sequence 2, Appli
c 32	23	2.3	23	9	US-09-966-264-21	Sequence 21, Appli
c 33	22	2.2	22	9	US-09-966-264-7	Sequence 7, Appli
c 34	22	2.2	22	9	US-09-966-264-12	Sequence 12, Appli
c 35	22	2.2	22	9	US-09-966-264-13	Sequence 13, Appli
c 36	22	2.2	22	9	US-09-966-264-29	Sequence 29, Appli
c 37	21	2.1	21	9	US-09-966-264-10	Sequence 10, Appli
c 38	20	2.0	20	9	US-09-966-264-9	Sequence 9, Appli
c 39	20	2.0	20	9	US-09-966-264-24	Sequence 24, Appli
c 40	20	2.0	36	15	US-10-149-736-65	Sequence 65, Appli
c 41	20	2.0	175	12	US-10-424-599-36603	Sequence 36603, A
c 42	20	2.0	459	12	US-10-424-599-53311	Sequence 53311, A
c 43	20	2.0	690	15	US-10-131-827-8211	Sequence 8211, Ap
c 44	20	2.0	734	12	US-10-424-599-45245	Sequence 45245, A
c 45	20	2.0	859	15	US-10-027-632-33432	Sequence 33432, A

## ALIGNMENTS

RESULT 1  
US-09-966-264-2  
; Sequence 2, Application US/09966264  
; Patent No. US2002009015A1  
; GENERAL INFORMATION:  
; APPLICANT: Barber, Elizabeth K  
; TITLE OF INVENTION: Gene Expression Control Element DNA  
; FILE REFERENCE: 89603465001  
; CURRENT APPLICATION NUMBER: US/09/966,264  
; CURRENT FILING DATE: 2001-09-28  
; PRIOR APPLICATION NUMBER: US 60/237,079  
; PRIOR FILING DATE: 2000-09-30  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 996  
; TYPE: DNA  
; ORGANISM: human  
; FEATURE:  
; NAME/KEY: exon  
; LOCATION: (1)..(996)  
; OTHER INFORMATION:  
; NAME/KEY: misc feature  
; LOCATION: (710)..(996)  
; OTHER INFORMATION: Nucleotides 710-996 are homologous to a portion of human dystroph  
; OTHER INFORMATION: in DNA in the region of exon 79 except that nucleotides 860-996 a  
; OTHER INFORMATION: re inverted in comparison to the orientation of the same sequence  
; OTHER INFORMATION: in the dystrophin DNA  
US-09-966-264-2

Query Match 100.0%; Score 996; DB 9; Length 996;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 996; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGGTTGATTGATGATAAAAAATGTCGTTAATACAGTAGAGAGTAAGTAATCAAT 60  
Db 1 GTGGTTGATTGATGATAAAAAATGTCGTTAATACAGTAGAGAGTAAGTAATCAAT 60

```

QY 61 CAATCACTCATAGCAAGTGGAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT 120
Db 61 CAATCACTCATAGCAAGTGGAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT 120
QY 121 AGAAATCTTGCTTATCTATGGAATCTTTTGTATATATTTACATGGGAACCTGAAT 180
Db 121 AGAAATCTTGCTTATCTATGGAATCTTTTGTATATATTTACATGGGAACCTGAAT 180
QY 181 GTAGCTTGACATTTTCCATGTAAACACCACTAGCTGATCCAACTAAAGCTGATCTA 240
Db 181 GTAGCTTGACATTTTCCATGTAAACACCACTAGCTGATCCAACTAAAGCTGATCTA 240
QY 241 ACAACACAGTGTATGCTTCAATTAATAGGCTTTGCTTCTCTGGAACCTGGTAAA 300
Db 241 ACAACACAGTGTATGCTTCAATTAATAGGCTTTGCTTCTCTGGAACCTGGTAAA 300
QY 301 AATCAAACTGTTGCTGTACACCTCGATGAGCTTCTGTTGTTCTTCCACCAAAATG 360
Db 301 AATCAAACTGTTGCTGTACACCTCGATGAGCTTCTGTTGTTCTTCCACCAAAATG 360
QY 361 GGAATGATTTCCCAATGGCAAAAGAAACAGAGTATGCTATCTATCTGCACTTTTGT 420
Db 361 GGAATGATTTCCCAATGGCAAAAGAAACAGAGTATGCTATCTATCTGCACTTTTGT 420
QY 421 AAGTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 480
Db 421 AAGTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 480
QY 481 GCAGATGATTTGGGAGAGCGATGAGTCCCTTAGTATCAGTCATGACAGATGAAGAGGA 540
Db 481 GCAGATGATTTGGGAGAGCGATGAGTCCCTTAGTATCAGTCATGACAGATGAAGAGGA 540
QY 541 GCAGATGATTTGGGAGAGCGATGAGTCCCTTAGTATCAGTCATGACAGATGAAGAGGA 600
Db 541 GCAGATGATTTGGGAGAGCGATGAGTCCCTTAGTATCAGTCATGACAGATGAAGAGGA 600
QY 601 AAGAGGATTTAGACAGTGAAGGTTTCAAGAAATAAATCTATATTTTGTGAGGAGTGT 660
Db 601 AAGAGGATTTAGACAGTGAAGGTTTCAAGAAATAAATCTATATTTTGTGAGGAGTGT 660
QY 661 GTATTATAGTGTAGATTTTCAAGTGTCTTCAAGTGTCTTCAAGTGTCTTCAAGTGT 720
Db 661 GTATTATAGTGTAGATTTTCAAGTGTCTTCAAGTGTCTTCAAGTGTCTTCAAGTGT 720
QY 721 GTTTTACAGTCTATGCAATTTTACAAAAGTTTATAAGAAACTACATGTAATAATCTTG 780
Db 721 GTTTTACAGTCTATGCAATTTTACAAAAGTTTATAAGAAACTACATGTAATAATCTTG 780
QY 781 ATAGCTAAATACCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 840
Db 781 ATAGCTAAATACCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 840
QY 841 TAACAGTTTATAAGAAAGATTTATAAGGAAAGAAATAACGCAATGGCAAGTGGTG 900
Db 841 TAACAGTTTATAAGAAAGATTTATAAGGAAAGAAATAACGCAATGGCAAGTGGTG 900
QY 901 AGCTGTGAATCAGGTGTGCAATTTATCAGGAACACCCCAAAACCAAGTGGGTAGA 960
Db 901 AGCTGTGAATCAGGTGTGCAATTTATCAGGAACACCCCAAAACCAAGTGGGTAGA 960
QY 961 AATAGCATGAGAGCGCTGTTGATGTTAAATTAAT 996
Db 961 AATAGCATGAGAGCGCTGTTGATGTTAAATTAAT 996

```

RESULT 2

```

US-10-149-736-40
; Sequence 40, Application US/10149736
; Publication No. US20030216332A1
; GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences

```

```

; FILE REFERENCE: UM-06968
; CURRENT APPLICATION NUMBER: US/10149,736
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: PCT/US01/31126
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,848
; PRIOR FILING DATE: 2000-10-06
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 40
; LENGTH: 5339
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-149-736-40

```

```

Query Match 41.2%; Score 410; DB 15; Length 5339;
Best Local Similarity 100.0%; Pred. No. 2.6e-189;
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGAGAGCGATGAGTC 509
Db 4491 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGAGAGCGATGAGTC 4550
QY 510 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGAAATAAATGTTTACAATCTCTGATT 569
Db 4551 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGAAATAAATGTTTACAATCTCTGATT 4610
QY 570 CCGCATGTTTATAATATTCATACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
Db 4611 CCGCATGTTTATAATATTCATACAAAGAGGATTAGACAGTAAGAGTTTACAAG 4670
QY 630 AATAAATCTATATTTTGTGAAGGTTAGTGTATTTACTAGTATTTCACTAGTATTTCT 689
Db 4671 AATAAATCTATATTTTGTGAAGGTTAGTGTATTTACTAGTATTTCACTAGTATTTCT 4730
QY 690 AAGTCTGTATTTGTTTAAACAATGGCAGGTTTACAGTCTATGCAATTTGACAAA 749
Db 4731 AAGTCTGTATTTGTTTAAACAATGGCAGGTTTACAGTCTATGCAATTTGACAAA 4790
QY 750 AAGTTATAGAAACTACATGTAATAATCTTGTAGTAAATAAATGTTGCAATTTCTTTATA 809
Db 4791 AAGTTATAGAAACTACATGTAATAATCTTGTAGTAAATAAATGTTGCAATTTCTTTATA 4850
QY 810 TGAACGCAATTTGGGTTGTTTAAAAAATTTATAACAGTTTATAAAGAAAGA 859
Db 4851 TGAACGCAATTTGGGTTGTTTAAAAAATTTATAACAGTTTATAAAGAAAGA 4900

```

RESULT 3

```

US-10-149-736-39
; Sequence 39, Application US/10149736
; Publication No. US20030216332A1
; GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; APPLICANT: Harper, Scott Q.
; TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences
; FILE REFERENCE: UM-06968
; CURRENT APPLICATION NUMBER: US/10149,736
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: PCT/US01/31126
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,848
; PRIOR FILING DATE: 2000-10-06
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 39
; LENGTH: 5417
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic

```



US-10-149-736-39

Query Match 41.2%; Score 410; DB 15; Length 5417;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-189; Indels 0; Gaps 0;  
 Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGCAGAGCGATGGAGTC 509  
 Db 4569 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGCAGAGCGATGGAGTC 4628  
 QY 510 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGATTAATGTTTACAACTCTCTGATT 569  
 Db 4629 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGATTAATGTTTACAACTCTCTGATT 4688  
 QY 570 CCCGATGGTCTTTTATATATTCATACAAAGAGGATTTAGACAGTAAGAGTTTACAAG 629  
 Db 4689 CCCGATGGTCTTTTATATATTCATACAAAGAGGATTTAGACAGTAAGAGTTTACAAG 4748  
 QY 630 AAATAAATCTATATTTTGTGAAGGTAGTGGTATTAATCTAGTATGATTTCTAGTATTTCT 689  
 Db 4749 AAATAAATCTATATTTTGTGAAGGTAGTGGTATTAATCTAGTATGATTTCTAGTATTTCT 4808  
 QY 690 AAGTCTGTTATTTGTTTAAACATGGCAGGTTTACACGCTCTATGCAATTTGACAAA 749  
 Db 4809 AAGTCTGTTATTTGTTTAAACATGGCAGGTTTACACGCTCTATGCAATTTGACAAA 4868  
 QY 750 AAGTTAAGAAAACACTACATGTAATCTTGATAGCTTAATAACTTGCATTTCTTTATA 809  
 Db 4869 AAGTTAAGAAAACACTACATGTAATCTTGATAGCTTAATAACTTGCATTTCTTTATA 4928  
 QY 810 TGGACGCAATTTTGGTGTGTTTAAATAATTTATAACAGTTATAAAGAAAAGA 859  
 Db 4929 TGGACGCAATTTTGGTGTGTTTAAATAATTTATAACAGTTATAAAGAAAAGA 4978

RESULT 4

US-10-149-736-41  
 ; Sequence 41, Application US/10149736  
 ; Publication No. US20030216332A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Chamberlain, Jeffrey S.  
 ; APPLICANT: Harper, Scott Q.  
 ; TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences  
 ; FILE REFERENCE: UM-06968  
 ; CURRENT APPLICATION NUMBER: US/10/149,736  
 ; CURRENT FILING DATE: 2002-06-17  
 ; PRIOR APPLICATION NUMBER: PCT/US01/31126  
 ; PRIOR FILING DATE: 2001-10-04  
 ; PRIOR APPLICATION NUMBER: 60/238,848  
 ; PRIOR FILING DATE: 2000-10-06  
 ; NUMBER OF SEQ ID NOS: 96  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 41  
 ; LENGTH: 5462  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Synthetic  
 US-10-149-736-41

Query Match 41.2%; Score 410; DB 15; Length 5462;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-189; Indels 0; Gaps 0;  
 Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGCAGAGCGATGGAGTC 509  
 Db 4614 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGCAGAGCGATGGAGTC 4673  
 QY 510 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGATTAATGTTTACAACTCTCTGATT 569  
 Db 4674 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGATTAATGTTTACAACTCTCTGATT 4733  
 QY 570 CCCGATGGTCTTTTATATATTCATACAAAGAGGATTTAGACAGTAAGAGTTTACAAG 629

Db 4734 CCCGATGGTCTTTTATATATTCATACAAAGAGGATTTACACAGTAAGAGTTTACAAG 4793  
 QY 630 AAATAAATCTATATTTTGTGAAGGTAGTGGTATTAATCTAGTATGATTTCTAGTATTTCT 689  
 Db 4794 AAATAAATCTATATTTTGTGAAGGTAGTGGTATTAATCTAGTATGATTTCTAGTATTTCT 4853  
 QY 690 AAGTCTGTTATTTGTTTAAACATGGCAGGTTTACACGCTCTATGCAATTTGTAACAAA 749  
 Db 4854 AAGTCTGTTATTTGTTTAAACATGGCAGGTTTACACGCTCTATGCAATTTGTAACAAA 4913  
 QY 750 AAGTTAAGAAAACACTACATGTAATCTTGATAGCTTAATAACTTGCATTTCTTTATA 809  
 Db 4914 AAGTTAAGAAAACACTACATGTAATCTTGATAGCTTAATAACTTGCATTTCTTTATA 4973  
 QY 810 TGGACGCAATTTTGGTGTGTTTAAATAATTTATAACAGTTATAAAGAAAAGA 859  
 Db 4974 TGGACGCAATTTTGGTGTGTTTAAATAATTTATAACAGTTATAAAGAAAAGA 5023

RESULT 5

US-10-149-736-42  
 ; Sequence 42, Application US/10149736  
 ; Publication No. US20030216332A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Chamberlain, Jeffrey S.  
 ; APPLICANT: Harper, Scott Q.  
 ; TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences  
 ; FILE REFERENCE: UM-06968  
 ; CURRENT APPLICATION NUMBER: US/10/149,736  
 ; CURRENT FILING DATE: 2002-06-17  
 ; PRIOR APPLICATION NUMBER: PCT/US01/31126  
 ; PRIOR FILING DATE: 2001-10-04  
 ; PRIOR APPLICATION NUMBER: 60/238,848  
 ; PRIOR FILING DATE: 2000-10-06  
 ; NUMBER OF SEQ ID NOS: 96  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 42  
 ; LENGTH: 8689  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Synthetic  
 US-10-149-736-42

Query Match 41.2%; Score 410; DB 15; Length 8689;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-189; Indels 0; Gaps 0;  
 Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGCAGAGCGATGGAGTC 509  
 Db 5985 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGCAGAGCGATGGAGTC 6044  
 QY 510 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGATTAATGTTTACAACTCTCTGATT 569  
 Db 6045 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGATTAATGTTTACAACTCTCTGATT 6104  
 QY 570 CCCGATGGTCTTTTATATATTCATACAAAGAGGATTTAGACAGTAAGAGTTTACAAG 629  
 Db 6105 CCCGATGGTCTTTTATATATTCATACAAAGAGGATTTAGACAGTAAGAGTTTACAAG 6164  
 QY 630 AAATAAATCTATATTTTGTGAAGGTAGTGGTATTAATCTAGTATGATTTCTAGTATTTCT 689  
 Db 6165 AAATAAATCTATATTTTGTGAAGGTAGTGGTATTAATCTAGTATGATTTCTAGTATTTCT 6224  
 QY 690 AAGTCTGTTATTTGTTTAAACATGGCAGGTTTACACGCTCTATGCAATTTGTAACAAA 749  
 Db 6225 AAGTCTGTTATTTGTTTAAACATGGCAGGTTTACACGCTCTATGCAATTTGTAACAAA 6284  
 QY 750 AAGTTAAGAAAACACTACATGTAATCTTGATAGCTTAATAACTTGCATTTCTTTATA 809  
 Db 6285 AAGTTAAGAAAACACTACATGTAATCTTGATAGCTTAATAACTTGCATTTCTTTATA 6344

QY 810 TGAACGCAATTTGGTGTGTTTAAATAATTTAAACAGTTATAAAGAAAGA 859  
DB 6345 TGAACGCAATTTGGTGTGTTTAAATAATTTAAACAGTTATAAAGAAAGA 6394

## RESULT 6

US-10-149-736-44  
; Sequence 44, Application US/10149736  
; Publication No. US20030216332A1  
; GENERAL INFORMATION:  
; APPLICANT: Chamberlain, Jeffrey S.  
; APPLICANT: Harper, Scott Q.  
; TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences  
; FILE REFERENCE: UM-06968  
; CURRENT APPLICATION NUMBER: US/10/149,736  
; CURRENT FILING DATE: 2002-06-17  
; PRIOR APPLICATION NUMBER: PCT/US01/31126  
; PRIOR FILING DATE: 2001-10-04  
; PRIOR APPLICATION NUMBER: 60/238,848  
; PRIOR FILING DATE: 2000-10-06  
; NUMBER OF SEQ ID NOS: 96  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 44  
; LENGTH: 11443  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
US-10-149-736-44

Query Match 41.2%; Score 410; DB 15; Length 11443;  
Best Local Similarity 100.0%; Pred. No. 2.7e-189;  
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 450 AGGACACAATGTAGGAAGTCTTTCCACATGGCAGATGATTTGGCAGAGCGATGGAGTC 509  
DB 8739 AGGACACAATGTAGGAAGTCTTTCCACATGGCAGATGATTTGGCAGAGCGATGGAGTC 8798  
QY 510 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGCAATAAATGTTTACAACCTCCTGATT 569  
DB 8799 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGCAATAAATGTTTACAACCTCCTGATT 8858  
QY 570 CCGCATGTTTATTAATAATTCATCAACAAGAGGATTAACAGAGTTTACAAG 629  
DB 8859 CCGCATGTTTATTAATAATTCATCAACAAGAGGATTAACAGAGTTTACAAG 8918  
QY 630 AATTAATCTATATTTTGTGAAGGAGTGTGTTTACTGATGTTTACTGATGTTTCT 689  
DB 8919 AATTAATCTATATTTTGTGAAGGAGTGTGTTTACTGATGTTTACTGATGTTTCT 8978  
QY 690 AAGTCTGTTTATTTGTTTAAACAATGGCAGTTTACACGTTTATGCAATTTGTAACAAA 749  
DB 8979 AAGTCTGTTTATTTGTTTAAACAATGGCAGTTTACACGTTTATGCAATTTGTAACAAA 9038  
QY 750 AGTTATAGAAACCTACATGTAATAATCTGTAGCTAAATAACTGCAATTTCTTTATA 809  
DB 9039 AGTTATAGAAACCTACATGTAATAATCTGTAGCTAAATAACTGCAATTTCTTTATA 9098  
QY 810 TGAACGCAATTTGGTGTGTTTAAATAATTTAAACAGTTATAAAGAAAGA 859  
DB 9099 TGAACGCAATTTGGTGTGTTTAAATAATTTAAACAGTTATAAAGAAAGA 9148

## RESULT 7

US-10-149-736-47  
; Sequence 47, Application US/10149736  
; Publication No. US20030216332A1  
; GENERAL INFORMATION:  
; APPLICANT: Chamberlain, Jeffrey S.  
; APPLICANT: Harper, Scott Q.  
; TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences  
; FILE REFERENCE: UM-06968  
; CURRENT APPLICATION NUMBER: US/10/149,736

; CURRENT FILING DATE: 2002-06-17  
; PRIOR APPLICATION NUMBER: PCT/US01/31126  
; PRIOR FILING DATE: 2001-10-04  
; PRIOR APPLICATION NUMBER: 60/238,848  
; PRIOR FILING DATE: 2000-10-06  
; NUMBER OF SEQ ID NOS: 96  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 47  
; LENGTH: 12057  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
US-10-149-736-47

Query Match 41.2%; Score 410; DB 15; Length 12057;  
Best Local Similarity 100.0%; Pred. No. 2.7e-189;  
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 450 AGGACACAATGTAGGAAGTCTTTCCACATGGCAGATGATTTGGCAGAGCGATGGAGTC 509  
DB 11214 AGGACACAATGTAGGAAGTCTTTCCACATGGCAGATGATTTGGCAGAGCGATGGAGTC 11273  
QY 510 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGCAATAAATGTTTACAACCTCCTGATT 569  
DB 11274 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGCAATAAATGTTTACAACCTCCTGATT 11333  
QY 570 CCGCATGTTTATTAATAATTCATCAACAAGAGGATTAACAGAGTTTACAAG 629  
DB 11334 CCGCATGTTTATTAATAATTCATCAACAAGAGGATTAACAGAGTTTACAAG 11393  
QY 630 AATTAATCTATATTTTGTGAAGGAGTGTGTTTACTGATGTTTACTGATGTTTCT 689  
DB 11394 AATTAATCTATATTTTGTGAAGGAGTGTGTTTACTGATGTTTACTGATGTTTCT 11453  
QY 690 AAGTCTGTTTATTTGTTTAAACAATGGCAGTTTACACGTTTATGCAATTTGTAACAAA 749  
DB 11454 AAGTCTGTTTATTTGTTTAAACAATGGCAGTTTACACGTTTATGCAATTTGTAACAAA 11513  
QY 750 AAGTTATAGAAACCTACATGTAATAATCTGTAGCTAAATAACTGCAATTTCTTTATA 809  
DB 11514 AAGTTATAGAAACCTACATGTAATAATCTGTAGCTAAATAACTGCAATTTCTTTATA 11573  
QY 810 TGAACGCAATTTGGTGTGTTTAAATAATTTAAACAGTTATAAAGAAAGA 859  
DB 11574 TGAACGCAATTTGGTGTGTTTAAATAATTTAAACAGTTATAAAGAAAGA 11623

## RESULT 8

US-09-782-378A-22  
; Sequence 22, Application US/09782378A  
; Patent No. US20020102731A1  
; GENERAL INFORMATION:  
; APPLICANT: Hearing, Patrick  
; APPLICANT: Bahou, Wadie  
; APPLICANT: Sandalon, Ziv  
; APPLICANT: Gratenko, Dmitri  
; TITLE OF INVENTION: Adenoviral Vectors  
; FILE REFERENCE: STONYB-04970  
; CURRENT APPLICATION NUMBER: US/09/782,378A  
; CURRENT FILING DATE: 2001-02-12  
; PRIOR APPLICATION NUMBER: 60/237,747  
; PRIOR FILING DATE: 2000-10-02  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 22  
; LENGTH: 13957  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-782-378A-22

Query Match 41.2%; Score 410; DB 9; Length 13957;  
Best Local Similarity 100.0%; Pred. No. 2.7e-189;

```
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 450 AGGACACAAATGAGGAAGTCTTTTCCACATGCGAGATGATTGGCGAGCGATGGAGTC 509
DB 11253 AGGACACAAATGAGGAAGTCTTTTCCACATGCGAGATGATTGGCGAGCGATGGAGTC 11312
QY 510 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGAAATTAATGTTTACAACTCCTGATT 569
DB 11313 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGAAATTAATGTTTACAACTCCTGATT 11372
QY 570 CCCGATGCTGTTTATAATATTTATACACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
DB 11373 CCCGATGCTGTTTATAATATTTATACACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11432
QY 630 AAATAAATCTATATTTTGTGAGGAGTGTGTTATATCTAGATTTTCAAGTTTCT 689
DB 11433 AAATAAATCTATATTTTGTGAGGAGTGTGTTATATCTAGATTTTCAAGTTTCT 11492
QY 690 AAGTCCTGTTATTTGTTTAACTATGCAATGCGAGTTTACACGCTCTATGCAATTTGACAAA 749
DB 11493 AAGTCCTGTTATTTGTTTAACTATGCAATGCGAGTTTACACGCTCTATGCAATTTGACAAA 11552
QY 750 AAGTTATAGAAAACCTACATGTAATCTTGATAGCTAAATAAATCTTGCCATTTCTTTATA 809
DB 11553 AAGTTATAGAAAACCTACATGTAATCTTGATAGCTAAATAAATCTTGCCATTTCTTTATA 11612
QY 810 TGGACGCAATTTGGGTTGTTTAAATAATTTATAACAGTTTATAAGAAAGA 859
DB 11613 TGGACGCAATTTGGGTTGTTTAAATAATTTATAACAGTTTATAAGAAAGA 11662
RESULT 9
US-09-880-107-2284
; Sequence 2284, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2284
; LENGTH: 13957
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 M18533
US-09-880-107-2284
Query Match 41.2%; Score 410; DB 9; Length 13957;
Best Local Similarity 100.0%; Pred. No. 2.7e-189;
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 450 AGGACACAAATGAGGAAGTCTTTTCCACATGCGAGATGATTGGCGAGCGATGGAGTC 509
DB 11253 AGGACACAAATGAGGAAGTCTTTTCCACATGCGAGATGATTGGCGAGCGATGGAGTC 11312
QY 510 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGAAATTAATGTTTACAACTCCTGATT 569
DB 11313 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGAAATTAATGTTTACAACTCCTGATT 11372
QY 570 CCCGATGCTGTTTATAATATTTATACACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
DB 11373 CCCGATGCTGTTTATAATATTTATACACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11432
QY 630 AAATAAATCTATATTTTGTGAGGAGTGTGTTATATCTAGATTTTCAAGTTTCT 689
DB 11433 AAATAAATCTATATTTTGTGAGGAGTGTGTTATATCTAGATTTTCAAGTTTCT 11492
QY 690 AAGTCCTGTTATTTGTTTAACTATGCAATGCGAGTTTACACGCTCTATGCAATTTGACAAA 749
DB 11493 AAGTCCTGTTATTTGTTTAACTATGCAATGCGAGTTTACACGCTCTATGCAATTTGACAAA 11552
QY 750 AAGTTATAGAAAACCTACATGTAATCTTGATAGCTAAATAAATCTTGCCATTTCTTTATA 809
DB 11553 AAGTTATAGAAAACCTACATGTAATCTTGATAGCTAAATAAATCTTGCCATTTCTTTATA 11612
QY 810 TGGACGCAATTTGGGTTGTTTAAATAATTTATAACAGTTTATAAGAAAGA 859
DB 11613 TGGACGCAATTTGGGTTGTTTAAATAATTTATAACAGTTTATAAGAAAGA 11662
```

```
QY 630 AAATAAATCTATATTTTGTGAGGAGTGTGTTATATCTAGATTTTCAAGTTTCT 689
DB 11433 AAATAAATCTATATTTTGTGAGGAGTGTGTTATATCTAGATTTTCAAGTTTCT 11492
QY 690 AAGTCCTGTTATTTGTTTAACTATGCAATGCGAGTTTACACGCTCTATGCAATTTGACAAA 749
DB 11493 AAGTCCTGTTATTTGTTTAACTATGCAATGCGAGTTTACACGCTCTATGCAATTTGACAAA 11552
QY 750 AAGTTATAGAAAACCTACATGTAATCTTGATAGCTAAATAAATCTTGCCATTTCTTTATA 809
DB 11553 AAGTTATAGAAAACCTACATGTAATCTTGATAGCTAAATAAATCTTGCCATTTCTTTATA 11612
QY 810 TGGACGCAATTTGGGTTGTTTAAATAATTTATAACAGTTTATAAGAAAGA 859
DB 11613 TGGACGCAATTTGGGTTGTTTAAATAATTTATAACAGTTTATAAGAAAGA 11662
RESULT 10
US-10-149-736-1
; Sequence 1, Application US/10149736
; Publication No. US20030216332A1
; GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; APPLICANT: Harper, Scott Q.
; TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences
; FILE REFERENCE: UM-06968
; CURRENT APPLICATION NUMBER: US/10/149,736
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: PCT/US01/31126
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,948
; PRIOR FILING DATE: 2000-10-06
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 13957
; TYPE: DNA
; ORGANISM: Homo sapiens
; OTHER INFORMATION:
US-10-149-736-1
Query Match 41.2%; Score 410; DB 15; Length 13957;
Best Local Similarity 100.0%; Pred. No. 2.7e-189;
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 450 AGGACACAAATGAGGAAGTCTTTTCCACATGCGAGATGATTGGCGAGCGATGGAGTC 509
DB 11253 AGGACACAAATGAGGAAGTCTTTTCCACATGCGAGATGATTGGCGAGCGATGGAGTC 11312
QY 510 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGAAATTAATGTTTACAACTCCTGATT 569
DB 11313 CTTAGTATCAGTCATGACAGATGAAGAGGAGCAGAAATTAATGTTTACAACTCCTGATT 11372
QY 570 CCCGATGCTGTTTATAATATTTATACACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
DB 11373 CCCGATGCTGTTTATAATATTTATACACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11432
QY 630 AAATAAATCTATATTTTGTGAGGAGTGTGTTATATCTAGATTTTCAAGTTTCT 689
DB 11433 AAATAAATCTATATTTTGTGAGGAGTGTGTTATATCTAGATTTTCAAGTTTCT 11492
QY 690 AAGTCCTGTTATTTGTTTAACTATGCAATGCGAGTTTACACGCTCTATGCAATTTGACAAA 749
DB 11493 AAGTCCTGTTATTTGTTTAACTATGCAATGCGAGTTTACACGCTCTATGCAATTTGACAAA 11552
QY 750 AAGTTATAGAAAACCTACATGTAATCTTGATAGCTAAATAAATCTTGCCATTTCTTTATA 809
DB 11553 AAGTTATAGAAAACCTACATGTAATCTTGATAGCTAAATAAATCTTGCCATTTCTTTATA 11612
QY 810 TGGACGCAATTTGGGTTGTTTAAATAATTTATAACAGTTTATAAGAAAGA 859
DB 11613 TGGACGCAATTTGGGTTGTTTAAATAATTTATAACAGTTTATAAGAAAGA 11662
```

```

RESULT 11
US-10-342-887-434
; Sequence 434, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; PRIOR FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 434
; LENGTH: 14069
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-434

Query Match 41.2%; Score 410; DB 12; Length 14069;
Best Local Similarity 100.0%; Pred. No. 2.7e-189;
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 450 AGGACACATGTAGGAAGTCTTTCCACATGCGAGATGATTTGGCGAGAGCGATGGAGTC 509
DB 11365 AGGACACATGTAGGAAGTCTTTCCACATGCGAGATGATTTGGCGAGAGCGATGGAGTC 11424

QY 510 CTTAGTATCAGTCATGACAGATGAAGAAGAGGAGTAAATGTTTACAACTCCTGATT 569
DB 11425 CTTAGTATCAGTCATGACAGATGAAGAAGAGGAGTAAATGTTTACAACTCCTGATT 11484

QY 570 CCGCATGTTTATAAATTCATACAAAGAGGATTTAGACAGTAAAGTTTACAAAG 629
DB 11485 CCGCATGTTTATAAATTCATACAAAGAGGATTTAGACAGTAAAGTTTACAAAG 11544

QY 630 AATAAATCTATATTTTGTGAAGGTAGTGGTATTATATCTGTAGATTTTCAAGTCTTCT 689
DB 11545 AATAAATCTATATTTTGTGAAGGTAGTGGTATTATATCTGTAGATTTTCAAGTCTTCT 11604

QY 690 AGTCTGTTATTTTGTGTACAAATGCGAGGTTTACACGCTATGCAATTTGTACAAA 749
DB 11605 AGTCTGTTATTTTGTGTACAAATGCGAGGTTTACACGCTATGCAATTTGTACAAA 11664

QY 750 AAGTTATAAGAAACTACATGTAATAATCTTTGATAGCTAAATAAATTTGCCATTTCTTTATA 809
DB 11665 AAGTTATAAGAAACTACATGTAATAATCTTTGATAGCTAAATAAATTTGCCATTTCTTTATA 11724

QY 810 TGAACGCATTTGGGTTGTTTAAAAATTTATACAGTTTATAAGAAAGA 859
DB 11725 TGAACGCATTTGGGTTGTTTAAAAATTTATACAGTTTATAAGAAAGA 11774

RESULT 12
US-10-342-887-981
; Sequence 981, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao

```

```

; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 981
; LENGTH: 14082
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-981

Query Match 41.2%; Score 410; DB 12; Length 14082;
Best Local Similarity 100.0%; Pred. No. 2.7e-189;
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 450 AGGACACATGTAGGAAGTCTTTCCACATGCGAGATGATTTGGCGAGAGCGATGGAGTC 509
DB 11378 AGGACACATGTAGGAAGTCTTTCCACATGCGAGATGATTTGGCGAGAGCGATGGAGTC 11437

QY 510 CTTAGTATCAGTCATGACAGATGAAGAAGAGGAGTAAATGTTTACAACTCCTGATT 569
DB 11438 CTTAGTATCAGTCATGACAGATGAAGAAGAGGAGTAAATGTTTACAACTCCTGATT 11497

QY 570 CCGCATGTTTATAAATTCATACAAAGAGGATTTAGACAGTAAAGTTTACAAAG 629
DB 11498 CCGCATGTTTATAAATTCATACAAAGAGGATTTAGACAGTAAAGTTTACAAAG 11557

QY 630 AATAAATCTATATTTTGTGAAGGTAGTGGTATTATATCTGTAGATTTTCAAGTCTTCT 689
DB 11558 AATAAATCTATATTTTGTGAAGGTAGTGGTATTATATCTGTAGATTTTCAAGTCTTCT 11617

QY 690 AAGTCTGTTATTTGTTTAAACAATGCGAGGTTTACACGCTATGCAATTTGTACAAA 749
DB 11618 AAGTCTGTTATTTGTTTAAACAATGCGAGGTTTACACGCTATGCAATTTGTACAAA 11677

QY 750 AAGTTATAAGAAACTACATGTAATAATCTTTGATAGCTAAATAAATTTGCCATTTCTTTATA 809
DB 11678 AAGTTATAAGAAACTACATGTAATAATCTTTGATAGCTAAATAAATTTGCCATTTCTTTATA 11737

QY 810 TGAACGCATTTGGGTTGTTTAAAAATTTATACAGTTTATAAGAAAGA 859
DB 11738 TGAACGCATTTGGGTTGTTTAAAAATTTATACAGTTTATAAGAAAGA 11787

RESULT 13
US-10-341-434-108
; Sequence 108, Application US/10341434
; Publication No. US20030215835A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies
; TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes
; FILE REFERENCE: 90 204 205 R1
; CURRENT APPLICATION NUMBER: US/10/341,434
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 60/348,154
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: US 60/348,119
; PRIOR FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 108
; LENGTH: 14082
; TYPE: DNA

```

```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (703)..(11388)
; OTHER INFORMATION:
US-10-341-434-108

Query Match      41.2%; Score 410; DB 15; Length 14082;
Best Local Similarity 100.0%; Pred. No. 2.7e-189;
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 450 AGGACACAAATGAGGAGTCTTTCCACATGCGAGATGATTGGCGAGCGATGGAGTC 509
DB 11378 AGGACACAAATGAGGAGTCTTTCCACATGCGAGATGATTGGCGAGCGATGGAGTC 11437
QY 510 CTTAGTATCAGTCATGACAGATGAAGAAGAGGACGAGATAAATGTTTACAACTCCTGATT 569
DB 11438 CTTAGTATCAGTCATGACAGATGAAGAAGAGGACGAGATAAATGTTTACAACTCCTGATT 11497
QY 570 CCCGATCGTTTTTATATATTCATACAAACAAAGAGGATAGACAGTAAGATTACAG 629
DB 11498 CCCGATCGTTTTTATATATTCATACAAACAAAGAGGATAGACAGTAAGATTACAG 11557
QY 630 AAATAAATCTATATTTTGTGAAGGTAGTGTGTTATTAATCTAGATTTCAGTAGTTTCT 689
DB 11558 AAATAAATCTATATTTTGTGAAGGTAGTGTGTTATTAATCTAGATTTCAGTAGTTTCT 11617
QY 690 AAGTCGTATTTGTTTGTAAATGCGAGTTTACACGCTATGCAATGTACAAA 749
DB 11618 AAGTCGTATTTGTTTGTAAATGCGAGTTTACACGCTATGCAATGTACAAA 11677
QY 750 AAGTTATGAAGAACTACATGTAATCTTGATAGCTAAATTAATCTGCAATTCCTTTATA 809
DB 11678 AAGTTATGAAGAACTACATGTAATCTTGATAGCTAAATTAATCTGCAATTCCTTTATA 11737
QY 810 TGAACGCAATTTGGTTGTTTAAATAATTTAACAAGTTTAAAGAAAGA 859
DB 11738 TGAACGCAATTTGGTTGTTTAAATAATTTAACAAGTTTAAAGAAAGA 11787

RESULT 14
US-10-149-736-38
; Sequence 38, Application US/10149736
; Publication No. US20030216332A1
; GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; TITLE OF INVENTION: Mini-Dystrophin Nucleic Acids and Peptide Sequences
; FILE REFERENCE: UM-06968
; CURRENT APPLICATION NUMBER: US/10/149,736
; PRIOR FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: PCT/US01/31126
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,848
; PRIOR FILING DATE: 2000-10-06
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 38
; LENGTH: 2691
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-149-736-38

Query Match      39.8%; Score 396; DB 15; Length 2691;
Best Local Similarity 100.0%; Pred. No. 1.7e-182;
Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 464 GAAGCTTTTCCACATGGCAGATGATTGGCGAGCGATGGAGTCCCTAGTATCAGTCA 523
DB 1 GAAGCTTTTCCACATGGCAGATGATTGGCGAGCGATGGAGTCCCTAGTATCAGTCA 60
QY 524 TGACAGATGAAGAAGAGAGCAGATAAATGTTTAACTCCTGATTCCCGCATGCTGTTT 583

```

```

DB 61 TGACAGATGAAGAAGAGAGCAGATAAATGTTTAACTCTCTGATTCCGCAATGGTTTTT 120
QY 584 ATAATATTTCATACAAAGAGGATTAGACAGTAAGAGTTTACAAAGAAATAAATCTATAT 643
DB 121 ATAATATTTCATACAAAGAGGATTAGACAGTAAGAGTTTACAAAGAAATAAATCTATAT 180
QY 644 TTTTGTGAAGGTAGTGGTATTATTAATCTAGATTTCAGTAGTTTCTAAGTCTGTTATTGT 703
DB 181 TTTTGTGAAGGTAGTGGTATTATTAATCTAGATTTCAGTAGTTTCTAAGTCTGTTATTGT 240
QY 704 TTTGTTAAACAATGGCAGGTTTACACGCTATGCAATTTGTCACAAAAAGCTTAAAGAAAA 763
DB 241 TTTGTTAAACAATGGCAGGTTTACACGCTATGCAATTTGTCACAAAAAGCTTAAAGAAAA 300
QY 764 CTACATGTAATCTTGATGATGCTAAATAAATCTGCAATTTCTTTATATGGAAGCAATTTG 823
DB 301 CTACATGTAATCTTGATGATGCTAAATAAATCTGCAATTTCTTTATATGGAAGCAATTTG 360
QY 824 GGTGTTTAAATAATTTATAACAGTTTAAAGAAAGA 859
DB 361 GGTGTTTAAATAATTTATAACAGTTTAAAGAAAGA 396

RESULT 15
US-09-796-692-3505/c
; Sequence 3505, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077,001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: PastSeq for Windows Version 3.0
; SEQ ID NO 3505
; LENGTH: 430
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-796-692-3505

Query Match      14.7%; Score 146; DB 9; Length 430;
Best Local Similarity 100.0%; Pred. No. 1.3e-60;
Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 851 AAAGAAAGATTATAAGGAAAAAGAAATAACGCAATGGACAAAGTGGTGAAGCTGTGAA 910

```

Db	296	AAAGGAAGATTATAAAGGAAAAGAAAATAACGCAATGGACAAGTGGTGAAGCTGTGAA	237
Qy	911	CTCAGGTGTGCACAAATTATCAGGACACCCCAAAACCAAGTGAGGTAGTAATAGCATGA	970
Db	236	CTCAGGTGTGCACAAATTATCAGGACACCCCAAAACCAAGTGAGGTAGTAATAGCATGA	177
Qy	971	GAAGCCGTGTTTGATGTTTAATT	996
Db	176	GAAGCCGTGTTTGATGTTTAATT	151

Search completed: April 6, 2004, 17:51:13  
Job time : 450.211 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2004, 11:42:46 ; Search time 472.067 Seconds  
(without alignments)  
8963.144 Million cell updates/sec

Title: US-09-966-264D-2  
Perfect score: 996  
Sequence: 1 gtgttgatgattagataaa.....gtgttgatgattagtaatt 996

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_29Jan04.\*  
1: Geneseqn1980s.\*  
2: Geneseqn1990s.\*  
3: Geneseqn2000s.\*  
4: Geneseqn2001as.\*  
5: Geneseqn2001bs.\*  
6: Geneseqn2002s.\*  
7: Geneseqn2003as.\*  
8: Geneseqn2003bs.\*  
9: Geneseqn2003cs.\*  
10: Geneseqn2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	996	100.0	996	ABK86497	Abk86497 Human Apo
2	996	100.0	1230	ABK86462	ABK86462 Human Apo
3	996	100.0	1234	ABK86463	ABK86463 Human Apo
4	410	41.2	2463	ABT33375	ABT33375 NOVX DNA
5	410	41.2	5339	ABK81998	ABK81998 DNA encod
6	410	41.2	5417	ABK81997	ABK81997 DNA encod
7	410	41.2	5462	ABK81999	ABK81999 DNA encod
8	410	41.2	6689	ABK82000	ABK82000 DNA encod
9	410	41.2	11443	ABK82002	ABK82002 DNA encod
10	410	41.2	12923	1 AN90338	AN90338 Sequence
11	410	41.2	13957	6 ABK81959	ABK81959 cDNA enco
12	410	41.2	13957	6 ABT10904	ABT10904 Human bre
13	410	41.2	13957	6 ABN95786	ABN95786 Gene #228
14	410	41.2	13957	6 ABK81996	ABK81996 Human dys
15	396	39.8	2691	6 ABK81996	ABK81996 Human dys
16	259	36.0	13977	6 ABK81996	ABK81996 Human bon
17	260	26.1	3163	3 AA48571	AA48571 A rod sho
18	260	26.1	3172	3 AA48570	AA48570 A rod sho
19	260	26.1	4075	3 AA248569	AA248569 A rod sho
20	260	26.1	4402	3 AA248568	AA248568 A rod sho
21	260	26.1	4402	3 AA248567	AA248567 A rod sho
22	251	25.2	3747	3 AA248566	AA248566 A rod sho
23	158	15.9	158	6 ABK86469	ABK86469 Human apo

C 24	146	14.7	200	6	ABK86468	ABK86468 Human dys
C 25	146	14.7	455	9	ADD32504	ADD32504 Human mit
C 26	146	14.7	2691	6	ABK81996	ABK81996 Human dys
C 27	146	14.7	8689	6	ABK82000	ABK82000 DNA encod
C 28	146	14.7	11443	6	ABK82002	ABK82002 DNA encod
C 29	146	14.7	13957	6	ABK81959	ABK81959 cDNA enco
C 30	146	14.7	13957	6	ABT10904	ABT10904 Human bre
C 31	146	14.7	13957	6	ABN95786	ABN95786 Gene #228
C 32	146	14.7	13957	6	ABK81996	ABK81996 Human dys
C 33	145	14.6	2005	7	ABT33376	ABT33376 NOVX DNA
C 34	137	13.8	137	6	ABK86496	ABK86496 Human Apo
C 35	126	12.7	13977	6	ABK86462	ABK86462 Human bon
C 36	94	9.4	1044	6	ABK86464	ABK86464 Mouse gen
C 37	91	9.1	13815	2	AAV18885	AAV18885 Mus muscu
C 38	91	9.1	13815	6	ABK81960	ABK81960 cDNA enco
C 39	91	9.1	13815	6	ABT199799	ABT199799 Mouse isc
C 40	91	9.1	19307	2	AA27558	AA27558 Shuttle v
C 41	60	6.0	60	6	ABK86471	ABK86471 Human apo
C 42	46	4.6	959	6	ABN74601	ABN74601 Bovine em
C 43	42	4.2	42	6	ABK86474	ABK86474 Human apo
C 44	42	4.2	13815	2	AAV18885	AAV18885 Mus muscu
C 45	42	4.2	13815	6	ABK81960	ABK81960 cDNA enco

ALIGNMENTS

RESULT 1  
ABK86497  
ID ABK86497 standard; cDNA; 996 BP.  
AC ABK86497;  
XX  
XX 27-AUG-2002 (first entry)  
XX Human Apo-dystrophin-4 cDNA.  
XX  
XX Human; ss; gene: apo-dystrophin-4; inversion sequence; gene therapy;  
XX protein truncation; muscular dystrophy; leukaemia.  
XX Homo sapiens.

Key Location/Qualifiers  
CDS 1..996  
/tag= a  
/product= "Apo-dystrophin-4 protein appearing as  
AA098739"  
FT FT /partial  
FT FT /note= "No start or stop codon shown"  
FT FT /transl\_except= (pos:7..9,aa:Xaa)  
FT FT /transl\_except= (pos:49..51,aa:Xaa)  
FT FT /transl\_except= (pos:187..189,aa:Xaa)  
FT FT /transl\_except= (pos:202..204,aa:Xaa)  
FT FT /transl\_except= (pos:217..219,aa:Xaa)  
FT FT /transl\_except= (pos:253..255,aa:Xaa)  
FT FT /transl\_except= (pos:265..270,aa:Xaa-Xaa)  
FT FT /transl\_except= (pos:394..396,aa:Xaa)  
FT FT /transl\_except= (pos:547..549,aa:Xaa)  
FT FT /transl\_except= (pos:565..567,aa:Xaa)  
FT FT /transl\_except= (pos:616..618,aa:Xaa)  
FT FT /transl\_except= (pos:649..651,aa:Xaa)  
FT FT /transl\_except= (pos:682..684,aa:Xaa)  
FT FT /transl\_except= (pos:709..711,aa:Xaa)  
FT FT /transl\_except= (pos:790..792,aa:Xaa)  
FT FT /transl\_except= (pos:841..843,aa:Xaa)  
FT FT /transl\_except= (pos:850..852,aa:Xaa)  
FT FT /transl\_except= (pos:880..882,aa:Xaa)  
FT FT /transl\_except= (pos:907..909,aa:Xaa)  
FT FT /transl\_except= (pos:952..954,aa:Xaa)  
FT FT /transl\_except= (pos:982..984,aa:Xaa)  
FT FT /transl\_except= (pos:988..990,aa:Xaa)  
FT FT /note= "Xaa= unknown, encoded by in frame stop codon"





/transl\_except= (pos:451..453,aa:Xaa)  
/transl\_except= (pos:487..489,aa:Xaa)  
/transl\_except= (pos:499..504,aa:Xaa-Xaa)  
/transl\_except= (pos:628..630,aa:Xaa)  
/transl\_except= (pos:781..783,aa:Xaa)  
/transl\_except= (pos:799..801,aa:Xaa)  
/transl\_except= (pos:850..852,aa:Xaa)  
/transl\_except= (pos:883..885,aa:Xaa)  
/transl\_except= (pos:916..920,aa:Xaa)  
/transl\_except= (pos:945..947,aa:Xaa)  
/transl\_except= (pos:1024..1026,aa:Xaa)  
/transl\_except= (pos:1075..1077,aa:Xaa)  
/transl\_except= (pos:1084..1086,aa:Xaa)  
/transl\_except= (pos:1114..1116,aa:Xaa)  
/transl\_except= (pos:1141..1143,aa:Xaa)  
/transl\_except= (pos:1186..1188,aa:Xaa)  
/transl\_except= (pos:1216..1218,aa:Xaa)  
/transl\_except= (pos:1222..1224,aa:Xaa)  
/note= "Xaa= unknown, encoded by in frame stop codon"

GB2368064-A.

24-APR-2002.

16-JAN-2001; 2001GB-00001124.

30-SEP-2000; 2000US-0237079P.

(IMCR ) IMPERIAL CANCER RES TECHNOLOGY LTD.  
(BARB/) BARBER E.

Barber E;

WPI: 2002-429042/46.

P-PSDB; AAU98729.

New human regulatory polynucleotide, useful for treating disorders associated with protein truncation, particularly muscular dystrophy, and related peptides and antibodies.

Disclosure; Fig 4; 222pp; English.

The invention relates to a polynucleotide (I) comprising, or consisting of, apo-dystrophin-4 inversion sequence appearing as ABK86496, or its functional equivalents (e.g. the apo-dystrophin-4 cDNA sequence appearing as ABK86497). Also included are polynucleotides that hybridize to either strand of (I), a vector containing (I), a cell containing (I) or the vector, proteins and peptides encoded by (I), a protein homologous with human dystrophin that is expressed on cell surfaces in vivo antibodies (Ab) specific for the protein and method of screening for leukemia cells by analyzing DNA for presence of (I) or by detecting presence of (II). The apo-dystrophin-4 inversion sequence is a regulatory element that controls expression (transcription and translation) of associated DNA, and may allow read-through of stop codons. The apo-dystrophin-4 inversion sequence is used in gene therapy of diseases associated with truncation of proteins, particularly muscular dystrophy and also leukaemia, but more generally (I) is a regulatory sequence used to control expression of any attached gene. Analysis of DNA for (I), or detection of proteins (II) encoded by (I), can be used to screen for leukaemic cells and related diseases. Antibodies raised against (II) can be used therapeutically, to inhibit (II) activity, also to detect (II) in screening assays. The present sequence is the full length cDNA sequence for human apo-dystrophin-4 containing a plurality of stop codons some of which may be read through due to the presence of (I) in the apo-dystrophin-4 gene

Sequence 1230 BP; 404 A; 189 C; 259 G; 378 T; 0 U; 0 Other;

Query Match 100.0%; Score 996; DB 6; Length 1230;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 996; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GTGGTTTGATGTAGTAAAGAAATGTCGTTATACAGTAGAGTAAGTAATCAAT 60  
|||||

Db	235	GTGGTTTGATGTAGTAAAGAAATGTCGTTATACAGTAGAGTAAGTAATCAAT	294
Qy	61	CAATCACTCATAGCCAGGTGGAAGAGATGTATCCCATCATGGAATATTCCTGTTCTGAT	120
Db	295	CAATCACTCATAGCCAGGTGGAAGAGATGTATCCCATCATGGAATATTCCTGTTCTGAT	354
Qy	121	AGAAATCTTGCTTATCTATGGAATCTTTTGATATATATTTACATTCGGGACCTGAAT	180
Db	355	AGAAATCTTGCTTATCTATGGAATCTTTTGATATATATTTACATTCGGGACCTGAAT	414
Qy	181	GTAGCTTGACATTTTCCATGTAAACACCCAGTAGCCTGATCCAAACATTAAGCTGATCTA	240
Db	415	GTAGCTTGACATTTTCCATGTAAACACCCAGTAGCCTGATCCAAACATTAAGCTGATCTA	474
Qy	241	AAAAACAAGTGAATGGCTTCATTAATAGGCTTTGCTTCTTCCTGGAAACCTGTTGAA	300
Db	475	AAAAACAAGTGAATGGCTTCATTAATAGGCTTTGCTTCTTCCTGGAAACCTGTTGAA	534
Qy	301	AATCAAAACCTTGTGTGTACACCTCGATCGAGTCTCTGTGTGTCTTCCACGAGAAATG	360
Db	535	AATCAAAACCTTGTGTGTGTACACCTCGATCGAGTCTCTGTGTGTCTTCCACGAGAAATG	594
Qy	361	GGGAATGATTTCCAAATGGCAAGAAACAGAGTGTATCTATCTCTGACCTTTTGTGA	420
Db	595	GGGAATGATTTCCAAATGGCAAGAAACAGAGTGTATCTATCTCTGACCTTTTGTGA	654
Qy	421	AGTCTGTCTTCT	480
Db	655	AGTCTGTCTTCT	714
Qy	481	GCAGATGATTTGGCGAGCGAGTGGAGTCTCTAGTATCAGTCAATGACAGATGAAGAGA	540
Db	715	GCAGATGATTTGGCGAGCGAGTGGAGTCTCTAGTATCAGTCAATGACAGATGAAGAGA	774
Qy	541	GCAGATGATTTGGCGAGCGAGTGGAGTCTCTAGTATCAGTCAATGACAGATGAAGAGA	600
Db	775	GCAGATGATTTGGCGAGCGAGTGGAGTCTCTAGTATCAGTCAATGACAGATGAAGAGA	834
Qy	601	AAGAGGATTTAGACAGTAAGAGTTTACAAGAAATTAATCTATATTTTGTGAAGGGTAGT	660
Db	835	AAGAGGATTTAGACAGTAAGAGTTTACAAGAAATTAATCTATATTTTGTGAAGGGTAGT	894
Qy	661	GTATTTACTGTAGATTTCAAGTCTCTAGTCTCTAGTCTCTAGTCTCTAGTCTCTAGTCT	720
Db	895	GTATTTACTGTAGATTTCAAGTCTCTAGTCTCTAGTCTCTAGTCTCTAGTCTCTAGTCT	954
Qy	721	GTATTTACTGTAGATTTCAAGTCTCTAGTCTCTAGTCTCTAGTCTCTAGTCTCTAGTCT	780
Db	955	GTATTTACTGTAGATTTCAAGTCTCTAGTCTCTAGTCTCTAGTCTCTAGTCTCTAGTCT	1014
Qy	781	ATAGCTAAATTAACCTTGCATTTCTTTATATGGAACGCAATTTGGGTTGTTTAAAAATTTA	840
Db	1015	ATAGCTAAATTAACCTTGCATTTCTTTATATGGAACGCAATTTGGGTTGTTTAAAAATTTA	1074
Qy	841	TACAGTTTAAAGAAAGTAATTAAGGAAAGAAAGAAATTAAGCAATGACAGATGGTG	900
Db	1075	TACAGTTTAAAGAAAGTAATTAAGGAAAGAAAGAAATTAAGCAATGACAGATGGTG	1134
Qy	901	AAGCTGTAACCTCAGGTGTGCACAAATTTATCAGGAACACCCCAAAACCAAGTGAAGTGA	960
Db	1135	AAGCTGTAACCTCAGGTGTGCACAAATTTATCAGGAACACCCCAAAACCAAGTGAAGTGA	1194
Qy	961	AATGACATGAGAGCCGCTGTTGATGTTAATTAATTT 996	
Db	1195	AATGACATGAGAGCCGCTGTTGATGTTAATTAATTT 1230	

RESULT 3

ID ABK86463 standard; cDNA; 1234 BP.

XX

AC ABK86463;

XX



KW parasitic infection; Alzheimer's disease; stroke; forensic biology;  
 KW immunogen; non-human transgenic animal; gene therapy; gene; ds.  
 XX Unidentified.  
 OS WO200281517-A2.  
 PN 17-OCT-2002.  
 PD 22-JAN-2002; 2002WO-US020264.  
 XX 19-JAN-2001; 2001US-0262892P.  
 PR 23-JAN-2001; 2001US-0263598P.  
 PR 25-JAN-2001; 2001US-0264117P.  
 PR 26-JAN-2001; 2001US-0264478P.  
 PR 30-JAN-2001; 2001US-0263351P.  
 PR 02-MAR-2001; 2001US-0272870P.  
 PR 14-MAR-2001; 2001US-0275927P.  
 PR 15-MAR-2001; 2001US-0275990P.  
 PR 15-MAR-2001; 2001US-0276449P.  
 PR 23-MAR-2001; 2001US-0277358P.  
 PR 29-MAR-2001; 2001US-0279857P.  
 PR 20-APR-2001; 2001US-0285140P.  
 PR 30-APR-2001; 2001US-0287484P.  
 PR 17-MAY-2001; 2001US-0291701P.  
 PR 08-JUN-2001; 2001US-0296960P.  
 PR 10-JUL-2001; 2001US-0304353P.  
 PR 12-JUL-2001; 2001US-0304886P.  
 PR 09-AUG-2001; 2001US-0311289P.  
 PR 13-AUG-2001; 2001US-0311975P.  
 PR 16-AUG-2001; 2001US-0312937P.  
 PR 18-OCT-2001; 2001US-0330227P.  
 PR 29-NOV-2001; 2001US-0334198P.  
 XX (CURA-) CURAGEN CORP.  
 PA Decristofaro MF, Padigaru M, Miller C, Tchernev V, Zhong H;  
 PI Zhong M, Anderson D, Ballinger R, Gerlach V, Spytek KA, Rastelli L;  
 PI Kekuda R, Guo X, Zerkusen B, Andrew D, Mezes P, Patturajan M;  
 PI Burgess CE, Eisen A, Wolenc A, Baumgartner J, Shinkets RA, Gusev V;  
 PI Vermet CAM, Taupier RJ, Pena C, Shenoy S, Li L, Casman S, Boldog F;  
 PI Fernandes B, Smithson G, Malyankar U, Tallon B, Liu X;  
 XX WPI; 2003-058504/05.  
 DR P-PSDB; ABJ37909.  
 XX New polypeptides, designated as NOVX, useful for diagnosing and treating  
 PT infections, neurological diseases, cancer, allergy, and bone,  
 PT immunological, skin, renal, brain, muscle and autoimmune disorders.  
 XX Claim 9; Page 171; 672pp; English.  
 XX The invention relates to a novel isolated polypeptide, designated NOVX  
 CC (NOV1 - 33), consisting of a mature form of one of 61 sequences, given in  
 CC the specification, or its variant, where amino acid residue(s) in the  
 CC variant differ from the mature form, provided that the variant differs in  
 CC not more than 15 % of the amino acids from the sequence of the mature  
 CC form. The NOVX polypeptides, nucleic acids encoding the polypeptides, and  
 CC an antibody to the polypeptides, are useful for treating or preventing a  
 CC NOVX-associated disorder in humans and for treating a syndrome associated  
 CC with a human disease (NOVX-associated disorder). NOVX polypeptides and  
 CC the encoding nucleic acids, are useful for determining the presence of or  
 CC predisposition to a disease associated with altered levels of NOVX  
 CC polypeptide and polynucleotide, by measuring the level of polypeptide  
 CC expression or the amount of nucleic acid from a mammal and comparing it  
 CC with another mammal not having or not predisposed to the disease. NOVX  
 CC polypeptide is also useful for identifying an agent that binds to NOVX  
 CC and a cell expressing NOVX is useful for identifying an agent that

CC modulates the expression or activity of NOVX. The antibodies and a  
 CC polypeptide having 95 % sequence identity to NOVX polypeptide are useful  
 CC for treating a pathological state in a mammal. The antibodies are also  
 CC useful for determining the presence or amount of NOVX in a sample. NOVX  
 CC polypeptides, polynucleotides and antibodies specific for the  
 CC polypeptides are useful for treating or preventing disorders or syndromes  
 CC including trauma, viral, bacterial, fungal, protozoal, and parasitic  
 CC infections. They can also treat disorders such as e.g., Alzheimer's  
 CC disease or a stroke. The NOVX encoding nucleic acids are useful for  
 CC expressing the NOVX proteins, to detect NOVX mRNA, or a genetic lesion in  
 CC a NOVX gene and to modulate NOVX activity. NOVX sequences are also useful  
 CC for identifying a cell or tissue type in a biological sample, to amplify  
 CC DNA sequences from very small biological samples such as tissues e.g.  
 CC hair or skin or body fluids in forensic biology and as primers and probes  
 CC for use in identifying and/or cloning NOVX homologues in other cell  
 CC types. The NOVX proteins are useful as an immunogen to generate  
 CC antibodies which are useful for diagnostically monitoring protein levels  
 CC and modulating NOVX activity. Cells comprising NOVX nucleic acids are  
 CC useful for producing non-human transgenic animals which are useful for  
 CC studying the function and/or activity of NOVX protein and for identifying  
 CC and/or evaluating modulators of NOVX protein activity. The NOVX nucleic  
 CC acids can be used in gene therapy. This polynucleotide sequence  
 CC represents a NOVX DNA sequence of the invention  
 XX  
 SQ Sequence 2463 BP; 691 A; 599 C; 580 G; 593 T; 0 U; 0 Other;  
 Query Match 41.2%; Score 410; DB 7; Length 2463;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-190;  
 Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 450 AGGACCAATGTAGGAAGTCTTTTCCACATGGCAGATGATTGGGCGAGCGATGAGTGC 509  
 DB 2009 AGGACCAATGTAGGAAGTCTTTTCCACATGGCAGATGATTGGGCGAGCGATGAGTGC 2068  
 QY 510 CTTAGTATCATGTCATGACAGATGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 569  
 DB 2069 CTTAGTATCATGTCATGACAGATGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2128  
 QY 570 CCGCATGGTGTATTAATAATTCATACAAAGAGAGATTAGACAGTAAAGAGTTTACAAG 629  
 DB 2129 CCGCATGGTGTATTAATAATTCATACAAAGAGAGATTAGACAGTAAAGAGTTTACAAG 2188  
 QY 630 AATAAATCATATATTTTGTGAAGGTAGTGGTATTATATCTGTAGATTTTCAGTAGTTTCT 689  
 DB 2189 AATAAATCATATATTTTGTGAAGGTAGTGGTATTATATCTGTAGATTTTCAGTAGTTTCT 2248  
 QY 690 AAGTCTGTATTCGTTTGTAAACAATGGCAGGTTTACACGCTCTATGCAATTCACAAA 749  
 DB 2249 AAGTCTGTATTCGTTTGTAAACAATGGCAGGTTTACACGCTCTATGCAATTCACAAA 2308  
 QY 750 AAGTTATAAGAAAACATACATGTAATAATCTTGTATAGCTAAATAACTTGCATTTCTTTATA 809  
 DB 2309 AAGTTATAAGAAAACATACATGTAATAATCTTGTATAGCTAAATAACTTGCATTTCTTTATA 2368  
 QY 810 TGGACCGCATTTGGGTGTTTAAATAATTTATAACAGTTTATAAAGAAAGA 859  
 DB 2369 TGGACCGCATTTGGGTGTTTAAATAATTTATAACAGTTTATAAAGAAAGA 2418  
 RESULT 5  
 ABK81998  
 ID ABK81998 standard; DNA; 5339 BP.  
 AC ABK81998;  
 XX  
 DT 13-AUG-2002 (first entry)  
 XX  
 DE DNA encoding mini-dystrophin protein deltaR2-R21.  
 XX  
 KW Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;  
 KW Duchenne's muscular dystrophy; DMD; dystrophin; ds.  
 XX  
 OS Homo sapiens.

```

OS Synthetic.
XX WO200229056-A2.
XX
XX
XX 11-APR-2002.
XX
XX 04-OCT-2001; 2001WO-US031126.
XX
XX 06-OCT-2000; 2000US-0238848P.
XX
XX (UNMI ) UNIV MICHIGAN.
XX
XX Chamberlain JS, Harper SQ;
XX
XX WPI; 2002-435334/46.
XX
XX
XX A composition for preparing therapeutic drugs, has a mini-dystrophin
XX peptide comprising a specific number of spectrin-like repeat domains, or
XX PT a nucleic acid sequence encoding the mini-dystrophin peptide.
XX
XX
XX Example 6; Fig 13; 145pp; English.
XX
XX
XX The invention describes a composition comprising a mini-dystrophin
XX peptide comprising a spectrin-like repeat domain, where the domain
XX comprises n spectrin-like repeats, and contains no more than n spectrin-
XX like repeats, where n is an even number between 4-24, or a nucleic acid
XX encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the
XX polynucleotide encoding it is useful as a medicament, for preparing a
XX drug for therapeutic application and in the preparation of a composition
XX for treatment of muscle disease, e.g. Duchenne's muscular dystrophy
XX (DMD). This sequence represents a mini-dystrophin sequence of the
XX invention
XX
XX Sequence 5339 BP; 1638 A; 1191 C; 1187 G; 1323 T; 0 U; 0 Other;
XX
XX
XX Query Match 41.2%; Score 410; DB 6; Length 5339;
XX Best Local Similarity 100.0%; Pred. No. 1.5e-190;
XX Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX
XX QY 450 AGGACACAATGTAGGAAGTCTTTTCCACATGGGCAGATGATTTGGGCAGAGCGATGGAGTC 509
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 4491 AGGACACAATGTAGGAAGTCTTTTCCACATGGGCAGATGATTTGGGCAGAGCGATGGAGTC 4550
XX
XX QY 510 CTTAGTATCATGTCATGACAGATGAAGAGGAGCGAATAAATGTTTTCACACTCCTGATT 569
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 4551 CTTAGTATCATGTCATGACAGATGAAGAGGAGCGAATAAATGTTTTCACACTCCTGATT 4610
XX
XX QY 570 CCGCATGGTITTTTATAATTCATACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 4611 CCGCATGGTITTTTATAATTCATACAAAGAGGATTAGACAGTAAGAGTTTACAAG 4670
XX
XX QY 630 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTACTGTAGATTTCAGTAGTTTCT 689
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 4671 AAATAAATCTATATTTTTGTGAAGGGTAGTGGTATTACTGTAGATTTCAGTAGTTTCT 4730
XX
XX QY 690 AAGTCTGTTATGTTTGTAAACAATGGCAGGTTTACAGTCTATGCAATTTGTACAAA 749
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 4731 AAGTCTGTTATGTTTGTAAACAATGGCAGGTTTACAGTCTATGCAATTTGTACAAA 4790
XX
XX QY 750 AAGTTATAAGAAAACTACATGTAATAATCTTGATAGTAAATAAATCTTGCCATTTCTTTATA 809
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 4791 AAGTTATAAGAAAACTACATGTAATAATCTTGATAGTAAATAAATCTTGCCATTTCTTTATA 4850
XX
XX QY 810 TGGAACCCATTTGGGTGTTTAAAAAATTTATAACAGTTATAAAGAAGA 859
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 4851 TGGAACCCATTTGGGTGTTTAAAAAATTTATAACAGTTATAAAGAAGA 4900
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX
XX
XX RESULT 6
XX ABK81997
XX ID ABK81997 standard; DNA; 5417 BP.
XX
XX AC ABK81997;

```

```
Db 4929 TGGACGCAATTTGGTCTTTAAATTTAATACAGTTATAAGAAAGA 4978
|||||
RESULT 7
ABK81999
ID ABK81999 standard; DNA; 5462 BP.
XX
AC ABK81999;
XX
DT 13-AUG-2002 (first entry)
XX
DE DNA encoding mini-dystrophin protein deltaR2-R21+H3.
XX
KW Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;
KW Duchenne's muscular dystrophy; DMD; dystrophin; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200229056-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-US031126.
XX
PR 06-OCT-2000; 2000US-0238948P.
XX
PA (UNMI ) UNIV MICHIGAN.
XX
PI Chamberlain JS, Harper SQ;
XX
DR WPI; 2002-435334/46.
XX
PT A composition for preparing therapeutic drugs, has a mini-dystrophin
PT peptide comprising a specific number of spectrin-like repeat domains, or
PT a nucleic acid sequence encoding the mini-dystrophin peptide.
XX
PS Disclosure; Fig 14; 145pp; English.
XX
CC The invention describes a composition comprising a mini-dystrophin
CC peptide comprising a spectrin-like repeat domain, where the domain
CC comprises n spectrin-like repeats, and contains no more than n spectrin-
CC like repeats, where n is an even number between 4-24, or a nucleic acid
CC encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the
CC polynucleotide encoding it is useful as a medicament, for preparing a
CC drug for therapeutic application and in the preparation of a composition
CC for treatment of muscle disease, e.g. Duchenne's muscular dystrophy
CC (DMD). This sequence represents a mini-dystrophin sequence of the
CC invention
XX
SQ Sequence 5462 BP; 1668 A; 1225 C; 1212 G; 1357 T; 0 U; 0 Other;
Query Match 41.2%; Score 410; DB 6; Length 5462;
Best Local Similarity 100.0%; Pred. No. 1.5e-190;
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 450 AGGACACAATGTAGGAAGTCTTTCCACATGGCAGATGATTTGGGACAGCGATGGAGTC 509
Db 4614 AGGACACAATGTAGGAAGTCTTTCCACATGGCAGATGATTTGGGACAGCGATGGAGTC 4673
QY 510 CTTAGTATCATGTCATGACAGATGAAGAGGAGGAGAGTAATAATGTTTACAACTCCTGATT 569
Db 4674 CTTAGTATCATGTCATGACAGATGAAGAGGAGGAGTAATAATGTTTACAACTCCTGATT 4733
QY 570 CCGCATGGTTTTATTAATTTATCATCAACAAGAGGAGGATAGACATGAGGTTTACAAG 629
Db 4734 CCGCATGGTTTTATTAATTTATCATCAACAAGAGGAGGATAGACATGAGGTTTACAAG 4793
QY 630 AAATAAATCTATATTTTGTGAAGGGTAGTGGTATTATATCTAGATTTTCAGTAGTTTCT 689
Db 4794 AAATAAATCTATATTTTGTGAAGGGTAGTGGTATTATATCTAGATTTTCAGTAGTTTCT 4853
```

---

```
QY 690 AAGTCTGTTATTGTTTGTAAACAATGGCAGGTTTTACAGTCTATGCAATTTCAAAA 749
Db 4854 AAGTCTGTTATTGTTTGTAAACAATGGCAGGTTTTACAGTCTATGCAATTTCAAAA 4913
QY 750 AGTTTATAAGAAAACACTACATGTAATCTTTGATAGCTAATTAACCTTCCCATTTCTTTATA 809
Db 4914 AAGTTTATAAGAAAACACTACATGTAATCTTTGATAGCTAATTAACCTTCCCATTTCTTTATA 4973
QY 810 TGGAACGCATTTTGGGTTTGTAAATAATTTTAAACAGTTTATAAAGAAAGA 859
Db 4974 TGGAACGCATTTTGGGTTTGTAAATAATTTTAAACAGTTTATAAAGAAAGA 5023

RESULT 8
ABK82000
ID ABK82000 standard; DNA; 8689 BP.
XX
AC ABK82000;
XX
DT 13-AUG-2002 (first entry)
XX
DE DNA encoding mini-dystrophin protein deltaH2-R19.
XX
KW Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;
KW Duchenne's muscular dystrophy; DMD; dystrophin; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200229056-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-US031126.
XX
PR 06-OCT-2000; 2000US-0238948P.
XX
PA (UNMI ) UNIV MICHIGAN.
XX
PI Chamberlain JS, Harper SQ;
XX
DR WPI; 2002-435334/46.
XX
PT A composition for preparing therapeutic drugs, has a mini-dystrophin
PT peptide comprising a specific number of spectrin-like repeat domains, or
PT a nucleic acid sequence encoding the mini-dystrophin peptide.
XX
PS Disclosure; Fig 15; 145pp; English.
XX
CC The invention describes a composition comprising a mini-dystrophin
CC peptide comprising a spectrin-like repeat domain, where the domain
CC comprises n spectrin-like repeats, and contains no more than n spectrin-
CC like repeats, where n is an even number between 4-24, or a nucleic acid
CC encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the
CC polynucleotide encoding it is useful as a medicament, for preparing a
CC drug for therapeutic application and in the preparation of a composition
CC for treatment of muscle disease, e.g. Duchenne's muscular dystrophy
CC (DMD). This sequence represents a mini-dystrophin sequence of the
CC invention
XX
SQ Sequence 8689 BP; 2721 A; 1804 C; 1861 G; 2303 T; 0 U; 0 Other;
Query Match 41.2%; Score 410; DB 6; Length 8689;
Best Local Similarity 100.0%; Pred. No. 1.5e-190;
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 450 AGGACACAATGTAGGAAGTCTTTCCACATGGCAGATGATTTGGGACAGCGATGGAGTC 509
Db 5985 AGGACACAATGTAGGAAGTCTTTCCACATGGCAGATGATTTGGGACAGCGATGGAGTC 6044
QY 510 CTTAGTATCATGTCATGACAGATGAAGAGGAGGAGAGTAATAATGTTTACAACTCCTGATT 569
Db 6045 CTTAGTATCATGTCATGACAGATGAAGAGGAGGAGTAATAATGTTTACAACTCCTGATT 6104
```

[illegible]

RESULT 9  
ABK82002  
ID ABK82002 standard: DNA: 11443 BP:

AC ABK82002;

XX  
DT 13-AUG-2002 (first entry)

DE DNA encoding mini-dystrophin protein deltaR9-R16.

XX Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;  
KW Duchenne's muscular dystrophy; DMD; dystrophin; ds.  
KW

AA	Homo sapiens.	Synthetic.
OS		
OS		

AA  
PN  
WO200229056-A2.

PD 11-APR-2002.

04-OCT-2001; 2001WO-US031126.

XX  
PR 06-OCT-2000: 2000US-0238848P.

XX (UTNMI ) UNIV MICHIGAN: PA

XX  
PI  
Chamberlain JS, Harper SO;XX  
DR WPI: 2002-435334/46.

XX A composition for preparing therapeutic drugs, has a mini-dystrophin  
PT peptide comprising a specific number of spectrin-like repeat domains, or  
PT a nucleic acid sequence encoding the mini-dystrophin peptide.

XX PS Disclosure: Fig 17: 145pp: English.

XX The invention describes a composition comprising a mini-dystrophin  
CC peptide comprising a spectrin-like repeat domain, where the domain  
CC comprises n spectrin-like repeats, and contains no more than n spectrin-  
CC like repeats, where n is an even number between 4-24, or a nucleic acid  
CC encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the  
CC polynucleotide encoding it is useful as a medicament, for preparing a  
CC drug for therapeutic application and in the preparation of a composition  
CC for treatment of muscle disease, e.g. Duchenne's muscular dystrophy  
CC (DMD). This sequence represents a mini-dystrophin sequence of the  
CC invention

Sequence 11443 BP; 3707 A; 2339 C; 2502 G; 2895 T; 0 U; 0 Other; ;

Query Match 41.2%; Score 410; DB 6; Length 11443;

Query Match 41.2%; score 410; DB 9, 10  
Best Local Similarity 100.0%; Pred. No. 1.5e-190;

	Matches	410;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	450	AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGACCGATGGAGTC	509							
Db	8739	AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGGCAGACCGATGGAGTC	8798							
Qy	510	CTTAGTATCAGTCATGACAGATGAAGAAGCAGAGCAGAAATAAATGTTTTTACAACTCCCTGATT	569							
Db	8799	CTTAGTATCAGTCATGACAGATGAAGAAGCAGAGCAGAAATAAATGTTTTTACAACTCCCTGATT	8858							
Qy	570	CCGCGATGGTTTTTAATAATTCATACAAACAAAGAGGATTAGACAGTAAGACAGTTTACAAG	629							
Db	8859	CCGCGATGGTTTTTAATAATTCATACAAACAAAGAGGATTAGACAGTAAGACAGTTTACAAG	8918							
Qy	630	AAATAAATCTATATTTTGTGAAGGGPAGTGGTATTATACGTAGATTTTCAGTAGTTTCT	689							
Db	8919	AAATAAATCTATATTTTGTGAAGGGPAGTGGTATTATACGTAGATTTTCAGTAGTTTCT	8978							
Qy	690	AAGTCTGTTATTGTTTGTAAACAATGGCAGGTTTTACAGCTCTATCGCAATTTGACAAAA	749							
Db	8979	AAGTCTGTTATTGTTTGTAAACAATGGCAGGTTTTACAGCTCTATCGCAATTTGACAAAA	9038							
Qy	750	AAGTTATAAGAAAACTACATGTAAAACTTTGATAGCTATAAATACTTCCCATTTCTTTATA	809							
Db	9039	AAGTTATAAGAAAACTACATGTAAAACTTTGATAGCTATAAATACTTCCCATTTCTTTATA	9098							
Qy	810	TGGAACGCATTTGGGTGGTTTAAAAATTTATAACAGTTATAAAGAAGA	859							
Db	9099	TGGAACGCATTTGGGTGGTTTAAAAATTTATAACAGTTATAAAGAAGA	9148							

RESULT 10

AAN90338  
ID AAN90338 standard; cDNA; 12923 BP.

AX  
AC  
AAN90338;

29-MAR-1992 (first entry)

XX  
DE  
Sequence of human muscular dystrophy (MD) cDNA.

xx  
Dystrophin; muscular dystrophy; probe; antibody; diagnosis; prenatal;  
kw  
heterozygote; gene therapy; genetic screening; foetal screening; ss.  
kw

XX  
OS Homo sapiens.

XX	Key.	Location/Qualifiers
PH	CDS	209. .12923
PT		/*tag= a

XX PN WO8906286-A.

XX  
PD  
13-JUL-1989

16-DEC-1988: 88WO-US004504.

22-DEC-1987. 87IIS-00136618.

XX  
PA (CHTT.-) CHILDRENS MED CENT.XX  
PT Kunkel J.M. Monaco A. Hoffman EP. Koenig M:

XX  
DP WPT: 1989-220587/30.

DR P-PSDB; AAP90373.  
YY

PT Muscular dystrophy gene - used for prepn. of probes, dyscytrophin  
PT polypeptide and antibodies for diagnosis and therapy of muscular  
PT dystrophy.

PS Disclosure; Fig 5; 68pp; English.

XX The inventors claim an MD probe comprising a purified ss NA SQ which  
CC hybridises to at least a part of the MD gene; pure dystrophin (DS)  
CC

CC	polypeptide, purified NA encoding DS and antibodies (Ab) to DS. The
CC	probes are equal to or stronger than 10 of one of 12 cDNA sequences
CC	deposited as ATCC 58666-57677. The MD gene is human, or a murine Dmd gene
XX	
SQ	Sequence 12923 BP; 4296 A; 2613 C; 2963 G; 3044 T; 0 U; 7 Other:
	Query Match            41.2%; Score 410; DB 1; Length 12923;
	Best Local Similarity 100.0%; Pred. No. 1.5e-190;
	Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	450 AGGACCAATGTAGGAAGTCTTTTCACATGGCAGCATGATTGGCGAGCGATGGAGTC 509
DB	11253 AGGACCAATGTAGGAAGTCTTTTCACATGGCAGCATGATTGGCGAGCGATGGAGTC 11312
QY	510 CTTTAGTATCAGTCATCACAGATGAAGAAGGCAGCAATAAATGTTTTACAACCTCCTGATT 569
DB	11313 CTTTAGTATCAGTCATGACAGATGAAGAAGGCAGCAATAAATGTTTTACAACCTCCTGATT 11372
QY	570 CCCGCATGGTTTTTAATAATTATTCATACAAAGAGGATTGACAGTAGAGTTTACAAG 629
DB	11373 CCCGCATGGTTTTTAATAATTATTCATACAAAGAGGATTGACAGTAGAGTTTACAAG 11432
QY	630 AAATAAATCTAATTTTTTGCGAAGGGTAGTGTTATATACTGTAGATTTTCAGTAGTTTCT 689
DB	11433 AAATAAATCTAATTTTTTGCGAAGGGTAGTGTTATATACTGTAGATTTTCAGTAGTTTCT 11492
QY	690 AAGTCGTGTTATGTTTGTGTTAAACAATGGCAGGTTTTACACGTCCTATGCAATTTGTTCAAAA 749
DB	11493 AAGTCGTGTTATGTTTGTGTTAAACAATGGCAGGTTTTACACGTCCTATGCAATTTGTTCAAAA 11552
QY	750 AAGTTATAAGAAACACATGTAAAAATCTTGATPAGCTAAATAACTTGGCATTTCCTTTATA 809
DB	11553 AAGTTATAAGAAACACATGTAAAAATCTTGATPAGCTAAATAACTTGGCATTTCCTTTATA 11612
QY	810 TGGAACCCATTTTGGGTGTTTAAAAATTTTATAACAGTTTATAAAGAAAGA 859
DB	11613 TGGAACCCATTTTGGGTGTTTAAAAATTTTATAACAGTTTATAAAGAAAGA 11662

RESULT 11	
ABK81959	
ID	ABK81959 standard; DNA; 13957 BP.
XX	
AC	ABK81959;
XX	
DT	13-AUG-2002 (first entry)
XX	
DE	cDNA encoding human dystrophin.
XX	
KW	Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;
KW	Duchenne's muscular dystrophy; DMD; dystrophin; human; gene; ds.
XX	
OS	Homo sapiens.
XX	
FN	WO200229056-A2.
XX	
PD	11-APR-2002.
XX	
PF	04-OCT-2001; 2001WO-US031126.
XX	
PR	06-OCT-2000; 2000US-0238848P.
XX	
PA	(UNMI ) UNIV MICHIGAN.
XX	
PI	Chamberlain JS, Harper SQ;
XX	
DR	WPI; 2002-435334/46.
XX	
PT	A composition for preparing therapeutic drugs, has a mini-dystrophin
PT	peptide comprising a specific number of spectrin-like repeat domains, or
PT	a nucleic acid sequence encoding the mini-dystrophin peptide.
XX	
PS	Example 2; Fig 1; 145pp; English.

XX	The invention describes a composition comprising a mini-dystrophin
CC	peptide comprising a spectrin-like repeat domain, where the domain
CC	comprises n spectrin-like repeats, and contains no more than n spectrin-
CC	like repeats, where n is an even number between 4-24, or a nucleic acid
CC	encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the
CC	polynucleotide encoding it is useful as a medicament, for preparing a
CC	drug for therapeutic application and in the preparation of a composition
CC	for treatment of muscle disease, e.g. Duchenne's muscular dystrophy
CC	(DMD). This sequence represents a human dystrophin polynucleotide
CC	sequence used in the creation of the mini-dystrophin peptides of the
CC	invention
XX	
SQ	Sequence 13957 BP; 4602 A; 2781 C; 3122 G; 3452 T; 0 U; 0 Other;
	Query Match 41.2%; Score 410; DB 6; Length 13957;
	Best Local Similarity 100.0%; Prad. No. 1.5e-190; Mismatches 0; Indels 0; Gaps 0;
	Matches 410; Conservative 0;
QY	450 AGGACACAATGTAGGAATCTTTTCCACATCGGCAGATGATTGGCGCAGCGATGGAGTC 509
DB	11253 AGGACACAATGTAGGAAGTCTTTTCCACATCGGCAGATGATTGGCGCAGCGATGGAGTC 11312
QY	510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAAATAAATGTTTTACACTCTCTGATT 569
DB	11313 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAAATAAATGTTTTACACTCTCTGATT 11372
QY	570 CCCGATCGTTTTTTATAATATTCATPACAAAGAGGATTAGACAGTAAGAGTTTACAAG 629
DB	11373 CCCGATCGTTTTTTATAATATTCATPACAAAGAGGATTAGACAGTAAGAGTTTACAAG 11432
QY	630 AAATAAATCTATATTTTGTAAGGAGTGTGTTATFATACHTGTAGATTTCAGTAGTTTCT 689
DB	11433 AAATAAATCTATATTTTGTAAGGAGTGTGTTATFATACHTGTAGATTTCAGTAGTTTCT 11492
QY	690 AAGTCTGTATTGTTTTGTTAACATGGCAGTTTTTACACGCTCATGCCAATGTGACAAA 749
DB	11493 AAGTCTGTATTGTTTTGTTAACATGGCAGTTTTTACACGCTCATGCCAATGTGACAAA 11552
QY	750 AAGTTATAGAAACATCAATGTAAAAATCTTGATAGCTAAATAAATTCGCTTTCTTTATA 809
DB	11553 AAGTTATAGAAACATCAATGTAAAAATCTTGATAGCTAAATAAATTCGCTTTCTTTATA 11612
QY	810 TGGAACGCGATTTTGGGTGTTTTTAAAAATTTTAAACAGTTATATAAGAAAGA 859
b	11613 TGGAACGCGATTTTGGGTGTTTTTAAAAATTTTAAACAGTTATATAAGAAAGA 11562

RESULT 12	
ABT10904	
ID	ABT10904 standard; cDNA; 13957 BP.
XX	
XX	ABT10904;
XX	
XX	04-DEC-2002 (first entry)
DT	
DT	
XX	
XX	Human breast cancer associated coding sequence SEQ ID NO: 1038.
DE	
XX	
XX	Human; breast specific gene; breast cancer; differential expression;
KW	cytostatic; gene therapy; gene; ss.
KW	
XX	
XX	Homo sapiens.
XX	
XX	WG200259271-A2.
PN	
PD	
PD	01-AUG-2002.
XX	
XX	
XX	25-JAN-2002; 2002WO-US002176.
FF	
XX	
PR	25-JAN-2001; 2001US-0263757P.
PR	25-APR-2001; 2001US-0286090P.
PR	23-MAY-2001; 2001US-0292517P.
XX	
XX	





11613 TGGACGCAATTTGGTGTGTTAAATTTATACAGTTATAAGAAAGA 11662

RESULT 14

ABSG69900 standard; DNA; 13957 BP.

ABSG69900;

21-NOV-2002 (first entry)

Human dystrophin gene.

Vector; adenovirus; adeno-associated; adenosine deaminase gene; receptor; adenosine deaminase deficiency; severe combined immune deficiency; P&H; beta-chain; haemoglobin gene; beta-thalassemia; sickle cell disease; low density lipoprotein gene; familial hypercholesterolaemia; hypoxanthine-guanine phosphoribosyltransferase; Lesch-Nyhan syndrome; phenylalanine hydroxylase gene; gene therapy; phenylketonuria; dystrophin gene; muscular dystrophy; cystic fibrosis; immunostimulant; human cystic fibrosis transmembrane conductance regulator gene; antianemic; antilipase; notropic; cytotaxic; dermatological; human; alpha-1-antitrypsin; lysosomal glucocerebrosidase; ADA; HPR; lysosomal arylsulphatase A; ornithine transcarbamylase; AKSA; OTC; NP; purin nucleoside phosphorylase; gene; ds.

Homo sapiens.

US2002102731-A1.

01-AUG-2002.

12-FEB-2001; 2001US-00782378.

02-OCT-2000; 2000US-0237747P.

(UUNY ) UNIV NEW YORK STATE RES FOUND.

Hearing P, Bahou WF, Sandalon Z, Gnatenko DV;

WPI; 2002-690619/74.

Producing vector, by introducing vector having nucleotide sequence, adenovirus inverted terminal repeats and packaging sequence, and adeno-associated virus terminal repeat, into cell, and culturing cell.

Disclosure; Page 122-128; 191pp; English.

The present invention relates to a new method of producing a vector. The method involves introducing recombinant vector having nucleotide sequence (NS) having 5' and 3' end, left and right inverted terminal repeats of adenovirus flanking NS, adenovirus packaging sequence linked to inverted terminal repeat, and adeno-associated virus terminal repeat linked to 3' end of NS, into cell expressing adenovirus early gene lacking from vector and culturing cell to produce another vector. The method is useful for generating vectors, especially mad vectors. The method is useful in transferring nucleotide sequences of interest into a cell, for gene transfer applications (e.g. gene therapy) in vitro, ex vivo and in vivo. The nucleotide sequences are useful for treating diseases associated with it, i.e. adenosine deaminase gene associated with adenosine deaminase deficiency with severe combined immune deficiency, beta-chain of haemoglobin gene associated with beta-thalassemia and sickle cell disease, receptor for low density lipoprotein gene associated with familial hypercholesterolaemia, hypoxanthine-guanine phosphoribosyltransferase associated with Lesch-Nyhan syndrome, phenylalanine hydroxylase (PAH) gene associated with phenylketonuria, dystrophin gene associated with muscular dystrophy, and human cystic fibrosis transmembrane conductance regulator gene associated with cystic fibrosis. The present nucleic acid sequence represents a human disease gene sequence that was used in the methods of the invention

Sequence 13957 BP; 4602 A; 2781 C; 3122 G; 3452 T; 0 U; 0 Other;

Query Match 41.2%; Score 410; DB 6; Length 13957;  
Best Local Similarity 100.0%; Pred. No. 1.5e-190;  
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 450 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGCAGAGCGATGAGTC 509

DB 11253 AGGACACAATGTAGGAAGTCTTTTCCACATGGCAGATGATTTGGCAGAGCGATGAGTC 11312

QY 510 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGATAAATGTTTACAACCTCCTGATT 569

DB 11313 CTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGATAAATGTTTACAACCTCCTGATT 11372

QY 570 CCGCATGGTTTTTAATTAATTCATACAAACAAGAGGATTAGACAGTAAGAGTTTACAAG 629

DB 11373 CCGCATGGTTTTTAATTAATTCATACAAACAAGAGGATTAGACAGTAAGAGTTTACAAG 11432

QY 630 AATAAATCTATATTTTCTGAAGGAGTGGTATTATATCTGTAGATTTTCAGTAGTTTCT 689

DB 11433 AATAAATCTATATTTTCTGAAGGAGTGGTATTATATCTGTAGATTTTCAGTAGTTTCT 11492

QY 690 AAGTCTGTTATTGTTTTGTTAAACAATGGCAGGTTTTTACACGCTATGCAATTGTACAAA 749

DB 11493 AAGTCTGTTATTGTTTTGTTAAACAATGGCAGGTTTTTACACGCTATGCAATTGTACAAA 11552

QY 750 AAGTTATAGAAACACATGTAATAATCTGTAGCTAATAACTTGGCCATTTCCTTTATA 809

DB 11553 AAGTTATAGAAACACATGTAATAATCTGTAGCTAATAACTTGGCCATTTCCTTTATA 11612

QY 810 TGGACGCAATTTGGGTTGTTTAAAAATTTATAACAGTTTATAAGAAAGA 859

DB 11613 TGGACGCAATTTGGGTTGTTTAAAAATTTATAACAGTTTATAAGAAAGA 11662

RESULT 15

ABK81996

ID ABK81996 standard; DNA; 2691 BP.

AC ABK81996;

DT 13-AUG-2002 (first entry)

Human dystrophin 3' untranslated region.

Mini-dystrophin peptide; spectrin-like repeat domain; muscle disease;

Duchenne's muscular dystrophy; DMD; dystrophin; human; ds.

OS Homo sapiens.

FN WO200229056-A2.

PD 11-APR-2002.

PF 04-OCT-2001; 2001WO-US031126.

PR 06-OCT-2000; 2000US-0238848P.

XX (UNMI ) UNIV MICHIGAN.

XX Chamberlain JS, Harper SO;

XX WPI; 2002-435334/46.

A composition for preparing therapeutic drugs, has a mini-dystrophin peptide comprising a specific number of spectrin-like repeat domains, or a nucleic acid sequence encoding the mini-dystrophin peptide.

Disclosure; Fig 10; 145pp; English.

The invention describes a composition comprising a mini-dystrophin peptide comprising a spectrin-like repeat domain, where the domain comprises n spectrin-like repeats, and contains no more than n spectrin-like repeats, where n is an even number between 4-24, or a nucleic acid

CC encoding a mini-dystrophin peptide. The mini-dystrophin peptide or the  
 CC polynucleotide encoding it is useful as a medicament, for preparing a  
 CC drug for therapeutic application and in the preparation of a composition  
 CC for treatment of muscle disease, e.g. Duchenne's muscular dystrophy  
 CC (DMD). This sequence represents a human dystrophin polynucleotide  
 CC sequence used in the creation of the mini-dystrophin peptides of the  
 CC invention  
 XX  
 SQ Sequence 2691 BP; 860 A; 448 C; 461 G; 922 T; 0 U; 0 Other;  
 Query Match 39.8%; Score 396; DB 6; Length 2691;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-183;  
 Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 464 GAAGTCCTTTCCACATGGCAGATGATTTGGGCGAGCGATGGAGTCCCTTAGTATCAGTCA 523  
 Db 1 GAAGTCCTTTCCACATGGCAGATGATTTGGGCGAGCGATGGAGTCCCTTAGTATCAGTCA 60  
 QY 524 TGACAGATGAAGAAGGAGGAGATGAATGTTTGGGCGAGCGATGGAGTCCCTTAGTATCAGTCA 583  
 Db 61 TGACAGATGAAGAAGGAGGAGATGAATGTTTGGGCGAGCGATGGAGTCCCTTAGTATCAGTCA 120  
 QY 584 ATAATATTCATACACAAAGAGGATTAGACAGTAAGAGTTTACAGAAATAAATCTATAT 643  
 Db 121 ATAATATTCATACACAAAGAGGATTAGACAGTAAGAGTTTACAGAAATAAATCTATAT 180  
 QY 644 TTTTGTGAAGGAGTGGTATTAATTAATCTAGTATGATTTTCTAGTCTCTGTTATTTGT 703  
 Db 181 TTTTGTGAAGGAGTGGTATTAATTAATCTAGTATGATTTTCTAGTCTCTGTTATTTGT 240  
 QY 704 TTTGTTAAACAATGGCAGGTTTACACGTCCTATGCAATTTGACAAAAAGTTTATAAGAAAA 763  
 Db 241 TTTGTTAAACAATGGCAGGTTTACACGTCCTATGCAATTTGACAAAAAGTTTATAAGAAAA 300  
 QY 764 CTACATGTAATATCTTGATAGCTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 823  
 Db 301 CTACATGTAATATCTTGATAGCTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 360  
 QY 824 GGTGTTTAAATAATTTATACAGTTTATTAAGAAAGA 859  
 Db 361 GGTGTTTAAATAATTTATACAGTTTATTAAGAAAGA 396

Search completed: April 6, 2004, 15:10:53  
 Job time : 474.067 secs